

	Matches	55;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
Qy	2	GGCGTGTACGGTGGAGGCCCTATATAGCAGAGTCGTTT	TAGTCAACCGTCAAAACCGTC	60						
Db	228	GGCGTGTACGGTGGAGGCTCTATATAGCAGAGTCGTTT	TAGTCAACCGTCAAGTCGCC	170						

RESULT 7					
BQ248646/c					
LOCUS	BQ248646	566 bp	mRNA	linear	EST 03-MAY-2002
DEFINITION	Tae25006C10F Tae25	Triticum aestivum cDNA clone	Tae25006C10F,	mrna	
	sequence.				
ACCESSION	BQ248646				
VERSION	BQ248646.1	GI:20444522			
KEYWORDS	EST.				
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				

SOURCE	ORGANISM	REFERENCE
bread wheat.	Triticum aestivum	1 (bases 1 to 566)
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae	
	; Triticeae; Triticum.	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 566)
Cloutier, S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >870 bp
Plate: 006 row: C column: 10
Seg primer: M13 Forward.

```

FEATURES
  source
    Location/Qualifiers
      1..566
        /organism="Triticum aestivum"
        /cultivar="Glenlea"
        /db_xref="taxon:4565"
        /clone="raE5008C10r"
        /clone_lib="raE25"
        /tissue_type="developing seeds"
        /dev_stage="25 days after anthesis"
        /lab_host="E. coli DH10B"
      /note="Vector: pCMV-Sport6.0 (Invitrogen Technologies);
        Site1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
        of cultivar Glenlea 25 days post-anthesis"
      136 a 138 c 143 g 149 t
BASE COUNT
ORIGIN

```

```

Query Match      71.1%; Score 52.6; DB 14; Length 566;
  est Local Similarity 93.2%; Pred. No. 5.2e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2  GCGGTGTACGGTGGGAGGCGCTATATAGACAGAGCTCGTTTAGTGACCCGCAACCGTC 60
Db      228  GCGGTGTACGGTGGGAGGCTATATAGACAGAGCTCGTTTAGTGACCCGCAACCGTC 170

```

RESULT_8	BQ248555/c
LOCUS	BQ248555
DEFINITION	bread wheat.
ACCESSION	BQ248555
VERSION	BQ248555.1
KEYWORDS	GI:20444431
SOURCE	EST.
	Triticum aestivum CDNA clone TaE25007E01F, mRNA linear EST 03-MAY-2002 BQ248555 575 bp mrna

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end). Average insert size is >870 bp
Plate: 007 row: E column: 01
Seq primer: M13 Forward.

FEATURES	Location/Qualifiers
source	1. .575

```

i.: 373
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE25007E01F"
/clone_lib="TaE25"
/tissue_type="developing seed"
/dev_stage="25 days after ant
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0
Site1: NotI; Site2: MluI; m
of cultivar Glenlea 25 days p
141 a 140 c 142 g 151 t
BASE COUNT
ORIGIN

```

Query Match	71.1%	Score 52.6;	DB 14;	Length 575;
Best Local Similarity	93.2%	Pred. No. 5.3e-09;		
Matches 55; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

Qy 2 GCGGTGTACGGTGGAGGCCATATTAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60
|||||
Db 230 GCGGTGTACGGTGGAGGCTATATTAAGCAGAGCTCGTTAGTGAACCGTCAAGATCGCC 172

RESULT	9
LOCUS	BQ248056/c
DEFINITION	BQ248056 TaE25013F06F TaE25 Triticum aestivum CDNA clone TaE25013F06F, mRNA sequence.
ACCESSION	BQ248056
VERSION	BQ248056.1
KEYWORDS	GI:20443932
SOURCE	EST.
ORGANISM	bread wheat. Triticum aestivum

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 593)
Cloutier, S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier

Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
; Triticeae; Triticum.

195 Datoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end). Average insert size is >870 bp
Plate: 013 row: F column: 06
Seq primer: M13 Forward.

```

FEATURES
  source
    1. 593
      Location/Qualifiers
        /organism="Triticum aestivum"
        /cultivar="Glenlea"
        /db_xref="taxon:4565"

```

```

/clone="Tae25013F06F"
/clone_lib="Tae25"
/tissue_type="developing seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/notes="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"
BASE COUNT 145 a 146 c 147 g 155 t
ORIGIN
Query Match 71.1%; Score 52.6; DB 14; Length 593;
Best Local Similarity 93.2%; Pred. No. 5.3e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 60
|||||
b 228 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 170
|||||

RESULT 10
B68191/c
LOCUS
DEFINITION B68191 610 bp DNA linear GSS 18-JUN-1998
CIT978SK-A-492C12.TV CIT978SK Homo sapiens genomic clone A-492C12,
DNA sequence.
ACCESSION B68191
VERSION B68191.1 GI:2666901
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 610)
AUTHORS Adams M.D., Rounsley S.D., Field C.E., Bass S., Linher K., Golden
K., Berry K., Granger D., Suh E., Wible C., Kim U.-J., Shizuya H.,
Simon M. and Venter J.C.
TITLE Use of a human BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL Unpublished (1997)
COMMENT Other GSSs: CIT978SK-A-492C12.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: 77
Class: BAC ends.
FEATURES
Location/Qualifiers
1..610
/organism="Homo sapiens"
/db_xref="GDB:726476"
/db_xref="taxon:9606"
/clone="A-492C12"
/clone_lib="CIT978SK"
/sex="Female"
/cell_type="Fibroblast"
/notes="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
Caltech Human BAC Library A"
BASE COUNT 146 a 159 c 165 g 140 t
ORIGIN
Query Match 71.1%; Score 52.6; DB 17; Length 610;
Best Local Similarity 93.2%; Pred. No. 5.4e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 60
|||||
b 288 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 230
|||||

/clone="Tae25013F06F"
/clone_lib="Tae25"
/tissue_type="developing seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/notes="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"
BASE COUNT 145 a 146 c 147 g 155 t
ORIGIN
Query Match 71.1%; Score 52.6; DB 14; Length 593;
Best Local Similarity 93.2%; Pred. No. 5.3e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 60
|||||
b 228 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 170
|||||

RESULT 11
BQ248255/c
LOCUS
DEFINITION BQ248255 625 bp mRNA linear EST 03-MAY-2002
Tae25011C08F Tae25 Triticum aestivum cDNA clone Tae25011C08F, mRNA
sequence.
ACCESSION BQ248255
VERSION BQ248255.1 GI:20444131
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 625)
AUTHORS Cloutier, S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >870 bp
Plate: 011 row: C column: 08
Seq primer: M13 Forward.
FEATURES
Location/Qualifiers
1..625
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="Tae25011C08F"
/clone_lib="Tae25"
/tissue_type="developing seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/notes="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"
BASE COUNT 155 a 152 c 154 g 164 t
ORIGIN
Query Match 71.1%; Score 52.6; DB 14; Length 625;
Best Local Similarity 93.2%; Pred. No. 5.4e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 60
|||||
Db 228 GCGGTGTCGGTGGGAGGCTATATAAGCAGAGCTGTTTACTGTAACCGTCAACCGTC 170
|||||

RESULT 12
BQ248467/c
LOCUS
DEFINITION BQ248467 625 bp mRNA linear EST 03-MAY-2002
Tae25008F09F Tae25 Triticum aestivum cDNA clone Tae25008F09F, mRNA
sequence.
ACCESSION BQ248467
VERSION BQ248467.1 GI:20444343
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 625)
AUTHORS Cloutier, S.
TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
JOURNAL Unpublished (2002)
COMMENT Contact: Dr. Sylvie Cloutier

```

Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@em.agr.ca
was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >870 bp
Plate: 008 row: F column: 09
Seq primer: M13 Forward.

FEATURES

source

1. .625
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaE25008F09F"
/clone_lib="TaE25"
/tissue_type="developing seeds"
/dev_stage="25 days after anthesis"
/lab_host="E. coli DH10B"
/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"

E COUNT

155 a 152 c 155 g 163 t

..IGIN

Query Match 71.1%; Score 52.6; DB 14; Length 625;
Best Local Similarity 93.2%; Pred. No. 5.4e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60

|||||
Db 228 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 170

RESULT 13

BQ248453/c

LOCUS BQ248453 630 bp mRNA linear EST 03-MAY-2002
DEFINITION TaE25008H04F TaE25 Triticum aestivum cDNA clone TaE25008H04F, mRNA
sequence.

ACCESSION BQ248453.1 GI:20444329

VERSION EST.

KEYWORDS bread wheat.

SOURCE

ORGANISM

Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

1. (bases 1 to 630)

Wheat functional genomics - Glenlea developing seeds cDNA libraries
Cloutier, S.
Unpublished (2002)

Contact: Dr. Sylvie Cloutier

Cereal Research Centre, Agriculture and Agri-food Canada

195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9

Tel: (204) 983-2340

Fax: (204) 983-4604

Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).

Average insert size is >870 bp

Plate: 008 row: H column: 04

Seq primer: M13 Forward.

Location/Qualifiers

1. .630

/organism="Triticum aestivum"

/cultivar="Glenlea"

/db_xref="taxon:4565"

/clone="TaE25008H04F"

/clone_lib="TaE25"

/tissue_type="developing seeds"

/dev_stage="25 days after anthesis"

/lab_host="E. coli DH10B"

/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);
Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"

BASE COUNT 156 a 154 c 155 g 165 t

ORIGIN

Query Match 71.1%; Score 52.6; DB 14; Length 630;

Best Local Similarity 93.2%; Pred. No. 5.5e-09;

Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60

|||||
Db 228 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 170

RESULT 14

BQ248173/c

LOCUS BQ248173 631 bp mRNA linear EST 03-MAY-2002
DEFINITION TaE25012C11F TaE25 Triticum aestivum cDNA clone TaE25012C11F, mRNA
sequence.

ACCESSION BQ248173.

VERSION BQ248173.1 GI:20444049

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.

1. (bases 1 to 631)

AUTHORS Cloutier, S.

Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)

Contact: Dr. Sylvie Cloutier

Cereal Research Centre, Agriculture and Agri-food Canada

195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9

Tel: (204) 983-2340

Fax: (204) 983-4604

Email: scloutier@em.agr.ca

was cloned directionally, not all sequences generated with reverse
primer were from the 5' end (same with forward primer and 3' end).

Average insert size is >870 bp

Plate: 012 row: C column: 11

Seq primer: M13 Forward.

Location/Qualifiers

1. .631

/organism="Triticum aestivum"

/cultivar="Glenlea"

/db_xref="taxon:4565"

/clone="TaE25012C11F"

/clone_lib="TaE25"

/tissue_type="developing seeds"

/dev_stage="25 days after anthesis"

/lab_host="E. coli DH10B"

/note="vector: pCMV-SPORT6.0 (Invitrogen Technologies);

Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
of cultivar Glenlea 25 days post-anthesis"

BASE COUNT 156 a 154 c 155 g 166 t

ORIGIN

Query Match 71.1%; Score 52.6; DB 14; Length 631;

Best Local Similarity 93.2%; Pred. No. 5.5e-09;

Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60

|||||
Db 228 GCGGTGACGGTGGAGGCTATATAAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 170

RESULT 15

BQ248939/c

LOCUS BQ248939 633 bp mRNA linear EST 03-MAY-2002
DEFINITION TaE25002C05F TaE25 Triticum aestivum cDNA clone TaE25002C05F, mRNA
sequence.

ACCESSION BQ248939
 VERSION BQ248939.1 GI:20444815
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM *Triticum aestivum*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; *Triticum*.
 REFERENCE 1 (bases 1 to 633)
 AUTHORS Cloutier, S.
 TITLE Wheat functional genomics - Glenlea developing seeds cDNA libraries
 JOURNAL Unpublished (2002)
 COMMENT Contact: Dr. Sylvie Cloutier
 Cereal Research Centre, Agriculture and Agri-food Canada
 195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
 Tel: (204) 983-2340
 Fax: (204) 983-4604
 Email: scloutier@em.agr.ca
 was cloned directionally, not all sequences generated with reverse
 primer were from the 5' end (same with forward primer and 3' end).
 Average insert size is >870 bp
 Plate: 002 row: C column: 05
 Seq primer: M13 Forward.

FEATURES
 source
 1..633
 /organism="Triticum aestivum"
 /cultivar="Glenlea"
 /db_xref="taxon:4565"
 /clone="TaE25002C05F"
 /clone_lib="TaE25"
 /tissue_type="developing seeds"
 /dev_stage="25 days after anthesis"
 /lab_host="E. coli DH10B"
 /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
 Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
 of cultivar Glenlea 25 days post-anthesis"
 BASE COUNT 157 a 154 c 154 g 168 t
 ORIGIN

Query Match 71.1%; Score 52.6; DB 14; Length 633;
 Best Local Similarity 93.2%; Pred. NO. 5.5e-09;
 Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 GCGGTGTACGGTGGGAGGCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60
 ||||||||||||||| ||||||||||||||||||||||||||||||| |||
 b 228 GCGGTGTACGGTGGGAGGCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 170

Search completed: March 6, 2003, 00:22:23
 Job time : 560.372 secs

Plate: 014 row: A column: 07
Seq primer: M13 Forward.
Location/Qualifiers
1..502
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="Tae25014A07F"
/clone_lib="Tae25"
/tissue_type="developing seeds"
/lab_host="E. coli DH10B"
/note="Vector: PCMV-SPORT6.0 (Invitrogen Technologies); Site.1: NotI; Site.2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 25 days post-anthesis"
121 a 128 c 122 g 131 t

BASE COUNT
ORIGIN

Query Match 71.1%; Score 52.6; DB 14; Length 502;
Best Local Similarity 93.2%; Pred. No. 5e-09;
Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GCGGTCTACGGTGGAGGCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60
|||||
DB 228 GCGGTCTACGGTGGAGGCTATATAGCAGAGCTCGTTAGTGAACCGTCAAGTCGCC 170
|||||

RESULT 6
BQ248278/c

LOCUS BQ248278
DEFINITION BQ248278.1 GI:20444154
ACCESSION BQ248278
VERSION
KEYWORDS EST
SOURCE bread wheat.
ORGANISM Triticum aestivum
linear EST 03-MAY-2002
TaE25011A03F TaE25 Triticum aestivum cDNA clone TaE25011A03F, mRNA sequence.

REFERENCE 1 (bases 1 to 535)
Cloutier, S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.ca
was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >870 bp
Plate: 011 row: A column: 03
Seq primer: M13 Forward.
Location/Qualifiers
1..535
/organism="Triticum aestivum"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="Tae25011A03F"
/clone_lib="Tae25"
/tissue_type="developing seeds"
/lab_host="E. coli DH10B"
/note="Vector: PCMV-SPORT6.0 (Invitrogen Technologies); Site.1: NotI; Site.2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 25 days post-anthesis"
130 a 133 c 132 g 139 t 1 others

BASE COUNT
ORIGIN

Query Match 71.1%; Score 52.6; DB 14; Length 535;
Best Local Similarity 93.2%; Pred. No. 5.1e-09;

```

FEATURES
  source
    Seq primer: T7.
    Location/Qualifiers
      1..657
        /organism="Danio rerio"
        /db_xref="taxon:7955"
        /clone="R2PD clone CHBop576F21214Q3"
        /clone_lib="zebrafish kidney cDNA random primed, R2PD
        library no: 576"
        /dev_stage="adult"
        /note="Organ: kidney; Vector: pBK-CMV"
BASE COUNT      163 a   165 c   162 g   165 t       2 others
ORIGIN

Query Match
  Best Local Similarity 71.9%; Score 53.2; DB 9; Length 657;
  Matches 58; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||
Db 369 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||

QY 62 AACCGC 67
    |||||
309 AGCGCG 304

FEATURES
  source
    Seq primer: T7.
    Location/Qualifiers
      1..703
        /organism="Glycine max"
        /cultivar="Harosoy 63"
        /db_xref="taxon:3847"
        /clone_lib="GmaxSC"
        /tissue_type="Seed coats"
        /lab_host="E. coli strain XL0LR"
        /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
        cDNA library was constructed from polyA+ enriched mRNA
        from green seed coats in mid to late developmental stage
        , average fresh weight 250 mg per seed. Traces of pod and
        embryo tissue also present. Complementary DNA was
        synthesized from mRNA using an XhoI-poly(dT)
        linker-primer. EcoRI adapters were ligated to the
        blunt-ended cDNA fragments and the products were digested
        with XhoI for directional cloning into lambda ZAP Express
        vector. This lambda library was amplified once using E.
        coli host strain XL1 Blue MRF+. Inserts were then
        subcloned by mass excision using ExAssist helper phage for
        conversion into phagemid vector pBK-CMV in E. coli host
        strain XL0LR."
BASE COUNT      199 a   147 c   148 g   203 t       6 others
ORIGIN

Query Match
  Best Local Similarity 71.9%; Score 53.2; DB 9; Length 657;
  Matches 58; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||
Db 369 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||

QY 62 AACCGC 67
    |||||
309 AGCGCG 304

FEATURES
  source
    Seq primer: T7.
    Location/Qualifiers
      1..703
        /organism="Glycine max"
        /cultivar="Harosoy 63"
        /db_xref="taxon:3847"
        /clone_lib="GmaxSC"
        /tissue_type="Seed coats"
        /lab_host="E. coli strain XL0LR"
        /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
        cDNA library was constructed from polyA+ enriched mRNA
        from green seed coats in mid to late developmental stage
        , average fresh weight 250 mg per seed. Traces of pod and
        embryo tissue also present. Complementary DNA was
        synthesized from mRNA using an XhoI-poly(dT)
        linker-primer. EcoRI adapters were ligated to the
        blunt-ended cDNA fragments and the products were digested
        with XhoI for directional cloning into lambda ZAP Express
        vector. This lambda library was amplified once using E.
        coli host strain XL1 Blue MRF+. Inserts were then
        subcloned by mass excision using ExAssist helper phage for
        conversion into phagemid vector pBK-CMV in E. coli host
        strain XL0LR."
BASE COUNT      199 a   147 c   148 g   203 t       6 others
ORIGIN

Query Match
  Best Local Similarity 71.9%; Score 53.2; DB 10; Length 703;
  Matches 58; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 2 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||
Db 535 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 61
    |||||

QY 62 AACCGC 67
    |||||
475 AGCGCG 470

FEATURES
  source
    Seq primer: T7.
    Location/Qualifiers
      1..329
        /organism="Homo sapiens"
        /db_xref="GDB:726891"
        /db_xref="taxon:9606"
        /clone="A-481E4"
        /clone_lib="CIT978SK"
        /sex="Female"
        /cell_type="Fibroblast"
        /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
        CalTech Human BAC Library A"
BASE COUNT      72 a   89 c   90 g   78 t

Query Match
  Best Local Similarity 71.1%; Score 52.6; DB 17; Length 329;
  Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 60
    |||||
Db 242 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 60
    |||||

RESULT 4
  LOCUS      BQ248969/c
  DEFINITION TAE25001G10F TAE25 Triticum aestivum cDNA clone TAE25001G10F, mRNA
  sequence.
  ACCESSION  BQ248969

```

```

FEATURES
  source
    Seq primer: T7.
    Location/Qualifiers
      1..329
        /organism="Homo sapiens"
        /db_xref="GDB:726891"
        /db_xref="taxon:9606"
        /clone="A-481E4"
        /clone_lib="CIT978SK"
        /sex="Female"
        /cell_type="Fibroblast"
        /note="Vector: pBAC108L; Site_1: HindIII; Site_2: HindIII;
        CalTech Human BAC Library A"
BASE COUNT      72 a   89 c   90 g   78 t

Query Match
  Best Local Similarity 71.1%; Score 52.6; DB 17; Length 329;
  Matches 55; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 60
    |||||
Db 242 GGCCTGTACGGTGGAGGCTATATAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTCA 60
    |||||

RESULT 4
  LOCUS      BQ248969/c
  DEFINITION TAE25001G10F TAE25 Triticum aestivum cDNA clone TAE25001G10F, mRNA
  sequence.
  ACCESSION  BQ248969

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 558.039 seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-3

Perfect score: 74

Sequence: 1 tggcgtgacgtggaggc.....accgtcaaccgcgggaagct 74

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Watch 100%

Listing first 45 summaries

Database :

```

EST:*
1: em_estba:*
2: em_esthma:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
C 1	53.2	71.9	657 9: AI815377	AI815377 B0283860
C 2	53.2	71.9	703 10: BE661741	BE661741 138c7 Gma
C 3	52.6	71.1	329 17: B51582	B51582 CIT978SK-A-
C 4	52.6	71.1	388 14: BQ248969	BQ248969 TAE25001G
C 5	52.6	71.1	502 14: BQ248026	BQ248026 TAE25014A
C 6	52.6	71.1	535 14: BQ248278	BQ248278 TAE25011A

C 7	52.6	71.1	566 14: BQ248646	BQ248646 TAE25006C
C 8	52.6	71.1	575 14: BQ248555	BQ248555 TAE25007E
C 9	52.6	71.1	593 14: BQ248056	BQ248056 TAE25013F
C 10	52.6	71.1	610 17: B68191	B68191 CIT978SK-A-
C 11	52.6	71.1	625 14: BQ248255	BQ248255 TAE25011C
C 12	52.6	71.1	625 14: BQ248467	BQ248467 TAE25008F
C 13	52.6	71.1	625 14: BQ248453	BQ248453 TAE25008H
C 14	52.6	71.1	630 14: BQ248453	BQ248453 TAE25012C
C 15	52.6	71.1	631 14: BQ248173	BQ248173 TAE25012C
C 16	52.6	71.1	633 14: BQ248939	BQ248939 TAE25002C
C 17	52.6	71.1	633 14: BQ248504	BQ248504 TAE25008B
C 18	52.6	71.1	637 14: BQ247936	BQ247936 TAE25015B
C 19	52.6	71.1	641 14: BQ248159	BQ248159 TAE25012E
C 20	52.6	71.1	641 14: BQ247761	BQ247761 TAE25042H
C 21	52.6	71.1	642 14: BQ247949	BQ247949 TAE25015A
C 22	52.6	71.1	644 14: BQ247963	BQ247963 TAE25014G
C 23	52.6	71.1	645 14: BQ247960	BQ247960 TAE25014H
C 24	52.6	71.1	646 14: BQ247762	BQ247762 TAE25042H
C 25	52.6	71.1	646 14: BQ248341	BQ248341 TAE25010C
C 26	52.6	71.1	647 14: BQ248311	BQ248311 TAE25010F
C 27	52.6	71.1	647 14: BQ248458	BQ248458 TAE25008G
C 28	52.6	71.1	648 14: BQ247997	BQ247997 TAE25014D
C 29	52.6	71.1	648 14: BQ248418	BQ248418 TAE25009C
C 30	52.6	71.1	648 14: BQ248913	BQ248913 TAE25002F
C 31	52.6	71.1	649 14: BQ247933	BQ247933 TAE25015B
C 32	52.6	71.1	649 14: BQ248366	BQ248366 TAE25009H
C 33	51	68.9	400 17: B63118	B63118 CIT978SK-A-
C 34	51	68.9	287 17: B63084	B63084 CIT978SK-A-
C 35	51	68.9	545 14: BQ248598	BQ248598 TAE25006H
C 36	50.8	68.6	550 14: BQ248595	BQ248595 TAE25006H
C 37	50.2	67.8	655 10: AW355220	AW355220 pnf-b.pk0
C 38	50.2	67.8	607 10: AV645401	AV645401 AV645401
C 39	50.2	67.8	617 10: AV645332	AV645332 AV645332
C 40	50.2	67.8	618 10: AV645383	AV645383 AV645383
C 41	50.2	67.8	619 10: AV645377	AV645377 AV645377
C 42	50.2	67.8	619 10: AV645379	AV645379 AV645379
C 43	50.2	67.8	619 10: AV645385	AV645385 AV645385
C 44	50.2	67.8	619 10: AV645389	AV645389 AV645389
C 45	50.2	67.8	620 10: AV645369	AV645369 AV645369
			620 10: AV645380	AV645380 AV645380

ALIGNMENTS

RESULT 1
AI815377/c
LOCUS B0283860 zebrafish kidney cDNA random primed, RZPD library no: 576
DEFINITION Danio rerio cDNA clone RZPD clone CHBOP576F21214Q3, mRNA sequence.
ACCESSION AI815377
VERSION AI815377.1 GI:5430923
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 657)
Look,A.I. and Holloway,M.
Zebrafish Kidney cDNA
Unpublished (1999)
Contact: Thomas Look
Dana-Farber Cancer Institute
Pediatric Oncology Department, Mayer Building-630.; 44 Binney Street
, Boston, MA 02115, USA
Tel: (617) 632-5826
Fax: (617) 632-6989
Email: thomas.look@dfci.harvard.edu
The clone was obtained from an RZPD array made from an adult kidney cDNA random primed library from Leonard Zon. RZPD library number 576. This clone is available from the Resource Centre/Primary Database of the German Human Genome Project: RZPD (Resource Centre/Primer Datenbank). <http://www.rzpd.edu>

QY 2 GCGGTGTACGGTGGAGGCCCTATATAACGAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 327 GCGGTGTACGGTGGAGGCCCTATATAACGAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385

RESULT 15
 AAX60047
 ID AAX60047 standard; DNA: 450 BP.
 AC AAX60047;
 XX
 XX
 DT 04-AUG-1999 (first entry)
 XX
 DE Minimal Human cytomegalovirus promoter.
 XX
 KW Transgenic mice; transgene; tet operator-linked gene; tetracycline;
 KW mouse-active transcriptional regulatory element; mutant Tet repressor;
 KW gene therapy; genetic disease; acquired disease; cancer; viral disease;
 KW vaccination; rheumatoid arthritis; hypopituitarism; wound healing;
 KW tissue regeneration; cancer; benign prostatic hypertrophy; hemophilia;
 KW erythrocytopenia; arteriosclerosis; liver disease; Alzheimer's disease;
 KW Parkinson's disease; human disease model; ds.
 Human cytomegalovirus.

XX
 PN US5912411-A.
 XX
 PD 15-JUN-1999.
 XX
 PF 07-JUN-1995; 93US-0487472.
 XX
 PR 07-JUN-1995; 95US-0487472.
 PR 14-JUN-1993; 93US-0076327.
 PR 14-JUN-1993; 93US-0076726.
 PR 14-JUN-1994; 94US-0260452.
 PR 01-JUL-1994; 94US-0270637.
 PR 15-JUL-1994; 94US-0275876.
 PR 03-FEB-1995; 95US-0383754.
 XX
 PA (UYHE-) UNIV HEIDELBERG.
 XX
 PI Bujard H, Gossen M;
 XX
 DR WPI; 1999-357232/30.
 XX
 PT Transgenic mice with inducible transgene activity useful for in
 PT vitro and in vivo protein production
 XX
 PS Disclosure; Column 71-72; 63pp; English.

The specification describes transgenic mice which have a transgene and a tet operator-linked gene integrated in the genome. The transgene comprises a mouse-active transcriptional regulatory element linked to a polynucleotide sequence that encodes a fusion protein which activates transcription of the tet operator-linked gene. The fusion protein comprises a mutated Tet repressor that binds a tet operator sequence in the presence of tetracycline (Tc) or a Tc analogue, linked to a polypeptide that activates transcription in eukaryotic cells. The transgenic system may be used for gene therapy to treat genes involved in genetic or acquired diseases. Gene therapy may be used to treat cancer, viral diseases, for vaccination, and to provide (Tc induced) regulated doses of a product (e.g. for the treatment or regulation of rheumatoid arthritis, hypopituitarism, wound healing and tissue regeneration, cancer, benign prostatic hypertrophy, hemophilia, erythrocytopenia, arteriosclerosis and liver disease, Alzheimer's disease, and Parkinson's disease). The system may also be used to produce proteins in vivo (e.g. using mammalian, yeast or fungal cells) or in vitro (e.g. transgenic farm animals), to produce animal models of human disease, or to produce a stable cell line for gene cloning. The present sequence is used to construct the transgenes of the invention.

Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 20; Length 450;
 Best Local Similarity 94.9%; Pred. No. 4.3e-12;
 Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 GCGGTGTACGGTGGAGGCCCTATATAACGAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 327 GCGGTGTACGGTGGAGGCCCTATATAACGAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385

Search completed: March 5, 2003, 22:49:57
 Job time : 73.2126 secs

CC (tTA) integrated at a predetermined location within chromosomal
CC DNA. The transgenic mice are useful for studying gene function and
CC expression. The method allows the development of knockout animals for
CC many genes of interest and allows for conditional inactivation or
CC modulation of expression of a gene in a mouse.

XX Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;
SQ

Query Match 73.2%; Score 54.2; DB 20; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GCGGTACGCTGGGGCCCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTACGCTGGGGCCCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 385

RESULT 13
ID AAX81722 standard; DNA; 450 BP.
AC AAX81722;
XX

27-AUG-1999 (first entry)

XX A human cytomegalovirus promoter sequence.

XX Tetracycline-controllable transactivator; tTA; transgenic mouse;
KW transgene; gene function; gene expression; knockout animal;
KW conditional inactivation; promoter; ds.
XX

OS Human cytomegalovirus.

XX US5922927-A.

PN 13-JUL-1999.

XX 21-JUL-1997; 97US-0897719.

XX 14-JUN-1994; 94US-0260452.

PR 14-JUN-1993; 93US-0076327.

PR 21-JUL-1997; 97US-0897719.

XX (BADI) BASF AG.

XX Bujard H, Gossen M, Salfeld JG, Voss JW;
WPI; 1999-404496/34.

XX Production of tetracycline-regulated transgenic mice, useful for
studying gene expression

PS Disclosure; Fig 7; 66pp; English.

XX The present sequence represents a human cytomegalovirus promoter.
CC It is used to create the transgenic animals of the invention. The
CC specification describes a method for producing a transgenic mouse
CC having a transgene encoding a tetracycline-controllable transactivator
CC (tTA) integrated at a predetermined location within chromosomal
CC DNA. The transgenic mice are useful for studying gene function and
CC expression. The method allows the development of knockout animals for
CC many genes of interest and allows for conditional inactivation or
CC modulation of expression of a gene in a mouse.

XX Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 20; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GCGGTACGCTGGGGCCCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTACGCTGGGGCCCTATATAGCAGAGCTCGTTAGTGAACCGTCAACCGTC 385

RESULT 14
ID AAX60046

XX AAX60046 standard; DNA; 450 BP.

AC AAX60046;

XX 04-AUG-1999 (first entry)

XX Minimal Human cytomegalovirus promoter.

XX Transgenic mice; transgene; tet operator-linked gene; tetracycline;
KW mouse-actve transcriptional regulatory element; mutant Tet repressor;
KW gene therapy; genetic disease; acquired disease; cancer; viral disease;
KW vaccination; rheumatoid arthritis; hypopituitarism; wound healing;
KW tissue regeneration; cancer; benign prostatic hypertrophy; hemophilia;
KW erythrocytopenia; arterosclerosis; liver disease; Alzheimer's disease;
KW Parkinson's disease; human disease model; ds.

XX Human cytomegalovirus.

XX US5912411-A.

XX 15-JUN-1999.

XX 07-JUN-1995; 93US-0487472.

XX 07-JUN-1995; 95US-0487472.

PR 14-JUN-1993; 93US-0076327.

PR 14-JUN-1993; 93US-0076726.

PR 14-JUN-1994; 94US-0260452.

PR 01-JUL-1994; 94US-0270637.

PR 15-JUL-1994; 94US-0275876.

PR 03-FEB-1995; 95US-0383754.

XX (UYHE-) UNIV HEIDELBERG.

XX Bujard H, Gossen M;

XX WPI; 1999-357232/30.

XX Transgenic mice with inducible transgene activity useful for in
vitro and in vivo protein production

PS Disclosure; Column 71-72; 63pp; English.

XX The specification describes transgenic mice which have a transgene and a
tet operator-linked gene integrated in the genome. The transgene
comprises a mouse-actve transcriptional regulatory element linked to a
polynucleotide sequence that encodes a fusion protein which activates
transcription of the tet operator-linked gene. The fusion protein
comprises a mutated Tet repressor that binds a tet operator sequence
in the presence of tetracycline (Tc) or a Tc analogue, linked to a
polypeptide that activates transcription in eukaryotic cells. The
transgenic system may be used for gene therapy to treat genes involved
in genetic or acquired diseases. Gene therapy may be used to treat
cancer, viral diseases, for vaccination, and to provide (Tc induced)
regulated doses of a product (e.g. for the treatment or regulation of
rheumatoid arthritis, hypopituitarism, wound healing and tissue
regeneration, cancer, benign prostatic hypertrophy, hemophilia,
erythrocytopenia, arterosclerosis and liver disease, Alzheimer's
disease, and Parkinson's disease). The system may also be used to
produce proteins in vivo (e.g. using mammalian, yeast or fungal cells)
or in vitro (e.g. transgenic farm animals), to produce animal models of
human disease, or to produce a stable cell line for gene cloning. The
present sequence is used to construct the transgenes of the invention.

XX Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 20; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

PI Bujard H, Gossen M;
XX WPI; 1998-541795/46.
XX
PT Tetracycline based regulation of gene expression - uses a
PT tetracycline operator sequence joined to a gene of interest, the
PT gene of interest being induced in the presence, but not absence of
PT the antibiotic
XX
XX Example 6; Columns 71-72; 63pp; English.
XX
CC The present sequence represents a Cytomegalovirus (CMV) minimal promoter
CC linked to 10 tet operator sequences. The specification describes a
CC method for regulating expression of a Tet (tetracycline) operator-linked
CC gene in a cell of a subject. The method comprises introducing into the
CC cell a nucleic acid encoding a fusion protein which inhibits
CC transcription in eukaryotic cells, the fusion protein comprising a
CC polypeptide which binds to a Tet operator sequence, operatively linked
CC to heterologous second polypeptide which inhibits transcription in
CC eukaryotic cells and modulating the concentration of a tetracycline
CC (analogue) in the subject. The method is used for the regulation of
CC gene expression system, using tetracycline (analogues). The system
CC enables a gene coupled to the system to be induced in the presence of
CC Tet and then stopped when Tet is removed.
...
SQ Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;
Query Match 73.2%; Score 54.2; DB 19; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 385
RESULT 11
AAV60081
ID AAV60081 standard; DNA; 450 BP.
XX
AC AAV60081;
XX
XX 04-DEC-1998 (first entry)
XX
DE Cytomegalovirus minimal promoter linked to 10 tet operator sequences.
XX
XX Tet repressor; tetracycline; regulation; expression;
KW Tet operator-linked gene; CMV minimal promoter; tet operator; ds.
XX
XX Synthetic.
XX Human cytomegalovirus.
XX
US5814618-A.
XX
PD 29-SEP-1998.
XX
XX 07-JUN-1995; 95US-0485978.
XX
PR 07-JUN-1995; 95US-0485978.
PR 14-JUN-1993; 93US-0076327.
PR 14-JUN-1993; 93US-0076726.
PR 14-JUN-1994; 94US-0260452.
PR 01-JUL-1994; 94US-0270637.
PR 15-JUL-1994; 94US-0275876.
PR 06-FEB-1995; 95US-0383754.
XX
XX (BADI) BASF AG.
XX (KNOL) KNOLL AG.
XX
PI Bujard H, Gossen M;
XX WPI; 1998-541795/46.
XX

PT Tetracycline based regulation of gene expression - uses a
PT tetracycline operator sequence joined to a gene of interest, the
PT gene of interest being induced in the presence, but not absence of
PT the antibiotic
XX
XX Disclosure; Columns 73-74; 63pp; English.
XX
CC The present sequence represents a Cytomegalovirus (CMV) minimal promoter
CC linked to 10 tet operator sequences. The specification describes a
CC method for regulating expression of a Tet (tetracycline) operator-linked
CC gene in a cell of a subject. The method comprises introducing into the
CC cell a nucleic acid encoding a fusion protein which inhibits
CC transcription in eukaryotic cells, the fusion protein comprising a
CC polypeptide which binds to a Tet operator sequence, operatively linked
CC to heterologous second polypeptide which inhibits transcription in
CC eukaryotic cells and modulating the concentration of a tetracycline
CC (analogue) in the subject. The method is used for the regulation of
CC gene expression system, using tetracycline (analogues). The system
CC enables a gene coupled to the system to be induced in the presence of
CC Tet and then stopped when Tet is removed.
XX
SQ Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;
Query Match 73.2%; Score 54.2; DB 19; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 385
RESULT 12
AAV81721
ID AAV81721 standard; DNA; 450 BP.
XX
AC AAV81721;
XX
XX 27-AUG-1999 (first entry)
XX
DE A human cytomegalovirus promoter sequence.
XX
XX Tetracycline-controllable transactivator; tTA; transgenic mouse;
KW transgene; gene function; gene expression; knockout animal;
KW conditional inactivation; promoter; ds.
XX
OS Human cytomegalovirus.
XX
XX US5922927-A.
XX
PD 13-JUL-1999.
XX
PF 21-JUL-1997; 97US-0897719.
XX
XX 14-JUN-1994; 94US-0260452.
PR 14-JUN-1993; 93US-0076327.
PR 21-JUL-1997; 97US-0897719.
XX
XX (BADI) BASF AG.
XX
XX Bujard H, Gossen M, Salfeld JG, Voss JW;
PI WPI; 1999-404496/34.
XX
XX Production of tetracycline-regulated transgenic mice, useful for
PT studying gene expression
PT
XX Example 2; Fig 6; 66pp; English.
XX
XX The present sequence represents a human cytomegalovirus promoter.
CC It is used to create the transgenic animals of the invention. The
CC specification describes a method for producing a transgenic mouse
CC having a transgene encoding a tetracycline-controllable transactivator

PR 01-JUL-1994; 94US-0270637.
 PR 15-JUL-1994; 94US-0275876.
 PR 03-FEB-1995; 95US-0383754.

XX (BUJA/) BUJARD H.
 PA (GOSS/) GOSSSEN M.

XX Bujard H, Gossen M;

XX WPI; 1996-087666/09.

XX New tetracycline-regulated transcription modulators - comprising
 PT fusion proteins which bind to tet operator sequences to activate or
 PT inhibit transcription

XX Disclosure; Page 73-74; 112pp; English.

XX Fusion proteins comprising a first polypeptide which binds to a tet
 CC operator sequence in the presence of tetracycline or a tetracycline
 CC analogue, operatively linked to a second polypeptide which either
 CC activates or inhibits transcription in eukaryotic cells may be used
 CC to activate or inhibit transcription. Such proteins may be used to
 CC regulate gene expression in cells and may be particularly useful for
 CC gene therapy and for expression of gene products in transgenic
 CC organisms. Induction of gene expression is rapid, efficient and
 CC strong, typically 1000-2000 fold. The inducing agent does not cause
 CC pleiotropic effects or cytotoxicity in eukaryotic cells.

CC Alternatively, a self regulating construct encoding a transactivator
 CC fusion protein can be created. Here, the fusion protein is
 CC operatively linked to a minimal promoter also comprising at least
 CC one tet operator sequence. This sequence is a CMV minimal promoter
 CC and ten tet operator sequences. Other promoters are described in
 CC AAT11356 and AAT11357.

XX Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 17; Length 450;

Best Local Similarity 94.9%; Pred. No. 4.3e-12;

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GCGGTGTACGTCGGAGCCCTATATAGCAGAGCTCGTTTACTGACCGTCAACCGTC 60
 Db 327 GCGGTGTACGTCGGAGCCCTATATAGCAGAGCTCGTTTACTGACCGTCAACCGTC 385

SSULT 9

AAT11356

ID AAT11356 standard; DNA; 450 BP.

XX AC AAT11356;

XX 07-JUL-1996 (first entry)

XX Minimal CMV promoter and ten tet operators.

XX Tet repressor; Herpes simplex virus; HSV; virion protein 16; VP16;
 KW fusion protein; gene expression; regulation; inhibition; activation;
 KW transcription; ds.

XX Transposon Tn-10/Cauliflower mosaic virus.

XX WO9601313-A1.

XX 18-JAN-1996.

XX 29-JUN-1995; 95WO-US08179.

XX 07-JUN-1995; 95US-0486814.

XX 01-JUL-1994; 94US-0270637.

XX 15-JUL-1994; 94US-0275876.

XX 03-FEB-1995; 95US-0383754.

XX (BUJA/) BUJARD H.

PA (GOSS/) GOSSSEN M.

XX Bujard H, Gossen M;

XX WPI; 1996-087666/09.

XX New tetracycline-regulated transcription modulators - comprising
 PT fusion proteins which bind to tet operator sequences to activate or
 PT inhibit transcription

XX Disclosure; Page 74; 112pp; English.

XX Fusion proteins comprising a first polypeptide which binds to a tet
 CC operator sequence in the presence of tetracycline or a tetracycline
 CC analogue, operatively linked to a second polypeptide which either
 CC activates or inhibits transcription in eukaryotic cells may be used
 CC to activate or inhibit transcription. Such proteins may be used to
 CC regulate gene expression in cells and may be particularly useful for
 CC gene therapy and for expression of gene products in transgenic
 CC organisms. Induction of gene expression is rapid, efficient and
 CC strong, typically 1000-2000 fold. The inducing agent does not cause
 CC pleiotropic effects or cytotoxicity in eukaryotic cells.

CC Alternatively, a self regulating construct encoding a transactivator
 CC fusion protein can be created. Here, the fusion protein is
 CC operatively linked to a minimal promoter also comprising at least
 CC one tet operator sequence. This sequence is a CMV minimal promoter
 CC and ten tet operator sequences. Other promoters are described in
 CC AAT11355 and AAT11357.

XX Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 17; Length 450;

Best Local Similarity 94.9%; Pred. No. 4.3e-12;

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GCGGTGTACGTCGGAGCCCTATATAGCAGAGCTCGTTTACTGACCGTCAACCGTC 60
 Db 327 GCGGTGTACGTCGGAGCCCTATATAGCAGAGCTCGTTTACTGACCGTCAACCGTC 385

RESULT 10

AAV60080

ID AAV60080 standard; DNA; 450 BP.

XX AC AAV60080;

XX 04-DEC-1998 (first entry)

XX Cytomegalovirus minimal promoter linked to 10 tet operator sequences.

XX Tet repressor; tetracycline; regulation; expression;

XX Tet operator-linked gene; CMV minimal promoter; tet operator; ds.

XX Synthetic.

XX Human cytomegalovirus.

XX US5814618-A.

XX 29-SEP-1998.

XX 07-JUN-1995; 95US-0485978.

XX 07-JUN-1995; 95US-0485978.

XX 14-JUN-1993; 93US-0076327.

XX 14-JUN-1993; 93US-0078726.

XX 14-JUN-1994; 94US-0260452.

XX 01-JUL-1994; 94US-0270637.

XX 15-JUL-1994; 94US-0275876.

XX 06-FEB-1995; 95US-0383754.

XX (BADI) BASF AG.

XX (KNOL) KNOLL AG.

```

QY 2 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 385

RESULT 6
AAQ76266
ID AAQ76266 standard; DNA; 450 BP.
XX
AC AAQ76266;
XX
DT 17-JUL-1995 (first entry)
XX
DE PhCMV*-1 promoter.
XX
KW tTA; transactivator; tetracycline-controllable transactivator;
KW conditional inactivation; homologous recombination; gene expression;
KW gene regulation; gene therapy; tetracycline-resistance; tetr;
KW transgenic animal; PhCMV*-1; promoter; tetr; CMV; ds.
XX
OS Human cytomegalovirus K12, Towne.
XX
XX Key Location/Qualifiers
XX mRNA 382..450
XX /*tag= a

XX WO9429442-A.
XX
PD 22-DEC-1994.
XX
PF 14-JUN-1994; 94WO-US06734.
XX
PR 14-JUN-1993; 93US-0076327.
XX
XX (BADI ) BASF AG.
XX
PI Bujard H, Gossen M, Salfeld JG, Voss JW;
XX WPI; 1995-036472/05.
XX
PT Regulatory systems using tetracycline-controllable transactivator
PT (tTA) - useful for conditional inactivation or modulation of
PT gene expression in a host cell or animal
XX
PS Disclosure; Page 53; 103pp; English.
XX
CC The minimal promoter PhCMV* spans the human CMV intermediate-
CC early promoter sequence from +75 to -53 (+1 being the first
CC nucleotide transcribed). TetO sequences were fused to
CC this core promoter to give the new promoters PhCMV*-1 (given
CC in AAQ76266) and PhCMV*-2 (AAQ76267). These promoters are used
CC to express tTA transactivator in host cells.
XX
SQ Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 16; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 385

RESULT 7
AAQ76267
ID AAQ76267 standard; DNA; 450 BP.
XX
AC AAQ76267;
XX
DT 17-JUL-1995 (first entry)
XX
DE PhCMV*-2 promoter.
XX

```

```

XX tTA; transactivator; tetracycline-controllable transactivator;
XX conditional inactivation; homologous recombination; gene expression;
XX gene regulation; gene therapy; tetracycline-resistance; tetr;
XX transgenic animal; PhCMV*-2; promoter; tetr; CMV; ds.
XX
OS Human cytomegalovirus Towne.
XX
XX Key Location/Qualifiers
XX mRNA 382..450
XX /*tag= a

XX WO9429442-A.
XX
PD 22-DEC-1994.
XX
PF 14-JUN-1994; 94WO-US06734.
XX
PR 14-JUN-1993; 93US-0076327.
XX
XX (BADI ) BASF AG.
XX
PI Bujard H, Gossen M, Salfeld JG, Voss JW;
XX WPI; 1995-036472/05.
XX
PT Regulatory systems using tetracycline-controllable transactivator
PT (tTA) - useful for conditional inactivation or modulation of
PT gene expression in a host cell or animal
XX
PS Disclosure; Page 53; 103pp; English.
XX
CC The minimal promoter PhCMV* spans the human CMV intermediate-
CC early promoter sequence from +75 to -53 (+1 being the first
CC nucleotide transcribed). TetO sequences were fused to
CC this core promoter to give the new promoters PhCMV*-1 (given
CC in AAQ76266) and PhCMV*-2 (AAQ76267). These promoters are used
CC to express tTA transactivator in host cells.
XX
SQ Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 16; Length 450;
Best Local Similarity 94.9%; Pred. No. 4.3e-12;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 60
DB 327 GCGGTGTACGGTGGGAGCGCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAACCGTC 385

RESULT 8
AAQ76268
ID AAQ76268 standard; DNA; 450 BP.
XX
AC AAQ76268;
XX
DT 07-JUL-1996 (first entry)
XX
DE Minimal CMV promoter and ten tet operators.
XX
KW Tet repressor; Herpes simplex virus; HSV; virion protein 16; VP16;
KW fusion protein; gene expression; regulation; inhibition; activation;
KW transcription; ds.
XX
OS Transposon Tn-10/Cauliflower mosaic virus.
XX
XX WO9601313-A1.
XX
PD 18-JAN-1996.
XX
PF 29-JUN-1995; 95WO-US08179.
XX
PR 07-JUN-1995; 95US-0486814.

```

XX Claim 12; Page 172; 241pp; English.

XX The invention relates to a construct which allows animals to be bred in

CC captivity but renders them infertile in the wild by allowing reversible

CC control over fertility and reproduction. The construct comprises a native

CC promoter, a blocking DNA sequence contoured for and designed to abrogate

CC a crucial gene's function or to cause its mis-expression, and a genetic

CC switch to regulate controlled expression/repression of the blocker/gene

CC knockout. The construct is useful for preventing embryogenesis or

CC gametogenesis in animals by stably transforming an animal cell with

CC the construct by microinjection, transfection or infection, where the

CC construct stably integrates into the genome by homologous recombination,

CC and implanting the cell into a host organism, where a whole animal

CC develops from the implanted cell. The present sequence is a repressible

CC promoter comprising tet responsive element (TRE) which is coupled to and

CC tightly regulates a minimal promoter region PminCMV from cytomegalovirus.

XX

Q Sequence 447 BP; 136 A; 104 C; 108 G; 99 T; 0 other;

Query Match 73.2%; Score 54.2; DB 22; Length 447;

Best Local Similarity 94.9%; Pred. No. 4.3e-12;

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGGAGCCCTATATAGCAGAGCTCGTTTACTGTAACCGTCAACCGTC 60

|||||

Db 371 GCGGTGTACGGTGGGAGCCCTATATAGCAGAGCTCGTTTACTGTAACCGTCAACCGTC 379

RESULT 4

AAT06869

ID AAT06869 standard; DNA; 450 BP.

XX AAT06869;

AC AAT06869;

XX 13-MAR-1996 (first entry)

XX PhCMV*-1 tetO construct.

DE

DE PhCMV*-1 tetO construct.

XX

XX Transactivator; tTA; tet operator; tetO; gene expression;

KW tetracycline-responsive promoter; PhCMV*-1; human cytomegalovirus;

KW transcription activation; eukaryotic cell; ds; ss.

XX

OS Synthetic.

XX

XX Key Location/Qualifiers

FT TATA_signal 347..353

FT /*tag= a

XX

XX US5464758-A.

XX

XX 07-NOV-1995.

XX

XX 14-JUN-1993; 93US-0076726.

XX

XX 14-JUN-1993; 93US-0076726.

XX

XX (BUJA/) BUJARD H.

PA (GOSS/) GOSSSEN M.

XX

XX Bujard H, Gossen M;

PI

XX WPI; 1995-392612/50.

XX

XX Polynucleotide encoding trans:activator fusion protein contg. tet

PT repressor - used to control expression of gene regulated by minimal

PT promoter linked to tet operon, and vectors and cells where gene

PT expression is regulated by tetracycline

XX

PS Disclosure; Fig 6; 37pp; English.

XX

XX The PhCMV*-1-tetO construct (AAT06869) is composed of tet operator

CC sequences separated by 95 bases from a human cytomegalovirus minimal

CC IE promoter. A transactivator, tTA or tTAs (AAT06867-68), is used to

CC regulate expression of a heterologous gene operably linked to

CC PhCMV*-1-tetO. On/off regulation of expression of the encoded

CC heterologous protein by host eukaryotic cells is provided by varying

CC the medium tetracycline conc.

XX

SQ Sequence 450 BP; 138 A; 101 C; 110 G; 101 T; 0 other;

Query Match 73.2%; Score 54.2; DB 16; Length 450;

Best Local Similarity 94.9%; Pred. No. 4.3e-12;

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGGAGCCCTATATAGCAGAGCTCGTTTACTGTAACCGTCAACCGTC 60

|||||

Db 327 GCGGTGTACGGTGGGAGCCCTATATAGCAGAGCTCGTTTACTGTAACCGTCAACCGTC 385

RESULT 5

AAT06870

ID AAT06870 standard; DNA; 450 BP.

XX AAT06870;

AC AAT06870;

XX 13-MAR-1996 (first entry)

XX PhCMV*-2 tetO construct.

DE

DE PhCMV*-2 tetO construct.

XX

XX Transactivator; tTA; tet operator; tetO; gene expression;

KW tetracycline-responsive promoter; PhCMV*-2; human cytomegalovirus;

KW transcription activation; eukaryotic cell; ds; ss.

XX

OS Synthetic.

XX

XX Key Location/Qualifiers

FT TATA_signal 347..353

FT /*tag= a

XX

XX US5464758-A.

XX

XX 07-NOV-1995.

XX

XX 14-JUN-1993; 93US-0076726.

XX

XX 14-JUN-1993; 93US-0076726.

XX

XX (BUJA/) BUJARD H.

PA (GOSS/) GOSSSEN M.

XX

XX Bujard H, Gossen M;

PI

XX WPI; 1995-392612/50.

XX

XX Polynucleotide encoding trans:activator fusion protein contg. tet

PT repressor - used to control expression of gene regulated by minimal

PT promoter linked to tet operon, and vectors and cells where gene

PT expression is regulated by tetracycline

XX

PS Disclosure; Fig 7; 37pp; English.

XX

XX The PhCMV*-2-tetO construct (AAT06870) is composed of tet operator

CC sequences separated by 76 bases from a human cytomegalovirus minimal

CC IE promoter. A transactivator, tTA or tTAs (AAT06867-68), is used to

CC regulate expression of a heterologous gene operably linked to

CC PhCMV*-2-tetO. On/off regulation of expression of the encoded

CC heterologous protein by host eukaryotic cells is provided by varying

CC the medium tetracycline conc.

XX

SQ Sequence 450 BP; 104 A; 109 C; 102 G; 135 T; 0 other;

Query Match 73.2%; Score 54.2; DB 16; Length 450;

Best Local Similarity 94.9%; Pred. No. 4.3e-12;

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

VERSION AR072128.1 GI:7223016
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Mice transgenic for a tetracycline-inducible transcriptional activator
JOURNAL Patent: US 5912411-A 8 15-JUN-1999;
FEATURES Location/Qualifiers
source 1..450
BASE COUNT 138 a 101 c 110 g 101 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
|||||
327 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385
|||||

RESULT 12
AR072129
LOCUS AR072129 450 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 9 from patent US 5912411.
ACCESSION AR072129
VERSION AR072129.1 GI:7223017
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Mice transgenic for a tetracycline-inducible transcriptional activator
JOURNAL Patent: US 5912411-A 9 15-JUN-1999;
FEATURES Location/Qualifiers
source 1..450
BASE COUNT 104 a 109 c 102 g 135 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
|||||
327 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385
|||||

RESULT 13
AR095982
LOCUS AR095982 450 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 8 from patent US 6004941.
ACCESSION AR095982
VERSION AR095982.1 GI:10024370
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Methods for regulating gene expression
JOURNAL Patent: US 6004941-A 8 21-DEC-1999;
FEATURES Location/Qualifiers
source 1..450
BASE COUNT 138 a 101 c 110 g 101 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
|||||
327 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385
|||||

RESULT 14
AR095983
LOCUS AR095983 450 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 9 from patent US 6004941.
ACCESSION AR095983
VERSION AR095983.1 GI:10024372
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Methods for regulating gene expression
JOURNAL Patent: US 6004941-A 9 21-DEC-1999;
FEATURES Location/Qualifiers
source 1..450
BASE COUNT 104 a 109 c 102 g 135 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
|||||
327 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385
|||||

RESULT 15
AR136642
LOCUS AR136642 450 bp DNA linear PAT 16-JUN-2001
DEFINITION Sequence 8 from patent US 6136954.
ACCESSION AR136642
VERSION AR136642.1 GI:14477314
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Tetracycline-inducible transcriptional activator fusion proteins
JOURNAL Patent: US 6136954-A 8 24-OCT-2000;
FEATURES Location/Qualifiers
source 1..450
BASE COUNT 138 a 101 c 110 g 101 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAAAACCGTC 60
|||||
327 GCGGTGTACGGTGGAGCCCTATATAAGCAGAGCTCGTTTGTAGTGAACCGTCAATCGCC 385
|||||

Search completed: March 5, 2003, 23:34:39
Job time : 408.715 secs

SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M., Salfeld,J.G. and Voss,J.W.
TITLE Mice transgenic for a tetracycline-controlled transcriptional activator
JOURNAL Patent: US 5859310-A 6 12-JAN-1999;
FEATURES Location/Qualifiers
 1..450
 /organism="unknown"
BASE COUNT 104 a 109 c 102 g 135 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 2 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAAACCGTC 60
|||||
Db 327 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAGATCGCC 385

RESULT 7
AR032151
LOCUS AR032151 450 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 8 from patent US 5866755.
ACCESSION AR032151
VERSION AR032151.1 GI:5946440
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Animals transgenic for a tetracycline-regulated transcriptional inhibitor
JOURNAL Patent: US 5866755-A 8 02-FEB-1999;
FEATURES Location/Qualifiers
 1..450
 /organism="unknown"
BASE COUNT 138 a 101 c 110 g 101 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 2 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAAACCGTC 60
|||||
Db 327 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAGATCGCC 385

RESULT 8
AR032152
LOCUS AR032152 450 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5866755.
ACCESSION AR032152
VERSION AR032152.1 GI:5946441
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Animals transgenic for a tetracycline-regulated transcriptional inhibitor
JOURNAL Patent: US 5866755-A 9 02-FEB-1999;
FEATURES Location/Qualifiers
 1..450
 /organism="unknown"
BASE COUNT 104 a 109 c 102 g 135 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 2 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAAACCGTC 60
|||||
Db 327 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAGATCGCC 385

RESULT 9
AR043816
LOCUS AR043816 450 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 8 from patent US 5814618.
ACCESSION AR043816
VERSION AR043816.1 GI:5964824
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Methods for regulating gene expression
JOURNAL Patent: US 5814618-A 8 29-SEP-1998;
FEATURES Location/Qualifiers
 1..450
 /organism="unknown"
BASE COUNT 138 a 101 c 110 g 101 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 2 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAAACCGTC 60
|||||
Db 327 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAGATCGCC 385

RESULT 10
AR043817
LOCUS AR043817 450 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 9 from patent US 5814618.
ACCESSION AR043817
VERSION AR043817.1 GI:5964825
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Methods for regulating gene expression
JOURNAL Patent: US 5814618-A 9 29-SEP-1998;
FEATURES Location/Qualifiers
 1..450
 /organism="unknown"
BASE COUNT 104 a 109 c 102 g 135 t
ORIGIN

Query Match 73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Y 2 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAAACCGTC 60
|||||
Db 327 GCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTACTGAACCGTCAAGATCGCC 385

RESULT 11
AR072128
LOCUS AR072128 450 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 8 from patent US 5912411.
ACCESSION AR072128

Gesellschaft fjr Biotechnologische Forschung mbH (GBF) (; DB)

```

FEATURES
  source
    Location/Qualifiers
      1..74
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT      19 a      17 c      23 g      15 t
ORIGIN

Query Match      100.0%; Score 74; DB 6; Length 74;
Best Local Similarity 100.0%; Pred. No. 7.5e-19;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCGGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 1 TGGCGGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60

QY 61 AAACCGGGAAGCT 74
      |||||||
Db 61 AAACCGGGAAGCT 74

RESULT 2
US
  I89986
  DEFINITION Sequence 1 from patent US 5723319.
  ACCESSION I89986
  VERSION I89986.1 GI:3409926
  KEYWORDS
  SOURCE Unknown.
  ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 432)
AUTHORS King,R.W., Barker,C.S. and Seeger,C.
TITLE Cultured cell line that inducibly expresses the hepatitis B virus genome, and uses thereof for screening antiviral substances
JOURNAL Patent: US 5723319-A 1 03-MAR-1998;
FEATURES
  source
    Location/Qualifiers
      1..432
        /organism="unknown"
BASE COUNT      135 a      94 c      105 g      98 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 432;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 315 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 373

ULT 3
J20199
LOCUS AR020199
DEFINITION Sequence 8 from patent US 5789156.
ACCESSION AR020199
VERSION AR020199.1 GI:3974814
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Tetracycline-regulated transcriptional inhibitors
JOURNAL Patent: US 5789156-A 8 04-AUG-1998;
FEATURES
  source
    Location/Qualifiers
      1..450
        /organism="unknown"
BASE COUNT      138 a      101 c      110 g      101 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 4
AR020200
LOCUS AR020200
DEFINITION Sequence 9 from patent US 5789156.
ACCESSION AR020200
VERSION AR020200.1 GI:3974815
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Tetracycline-regulated transcriptional inhibitors
JOURNAL Patent: US 5789156-A 9 04-AUG-1998;
FEATURES
  source
    Location/Qualifiers
      1..450
        /organism="unknown"
BASE COUNT      104 a      109 c      102 g      135 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 5
AR029414
LOCUS AR029414
DEFINITION Sequence 5 from patent US 5859310.
ACCESSION AR029414
VERSION AR029414.1 GI:5941387
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H., Gossen,M., Salfeld,J.G. and Voss,J.W.
TITLE Mice transgenic for a tetracycline-controlled transcriptional activator
JOURNAL Patent: US 5859310-A 5 12-JAN-1999;
FEATURES
  source
    Location/Qualifiers
      1..450
        /organism="unknown"
BASE COUNT      138 a      101 c      110 g      101 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 6
AR029415
LOCUS AR029415
DEFINITION Sequence 6 from patent US 5859310.
ACCESSION AR029415
VERSION AR029415.1 GI:5941388
KEYWORDS
  
```

```

Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 4
AR020200
LOCUS AR020200
DEFINITION Sequence 9 from patent US 5789156.
ACCESSION AR020200
VERSION AR020200.1 GI:3974815
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H. and Gossen,M.
TITLE Tetracycline-regulated transcriptional inhibitors
JOURNAL Patent: US 5789156-A 9 04-AUG-1998;
FEATURES
  source
    Location/Qualifiers
      1..450
        /organism="unknown"
BASE COUNT      104 a      109 c      102 g      135 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 5
AR029414
LOCUS AR029414
DEFINITION Sequence 5 from patent US 5859310.
ACCESSION AR029414
VERSION AR029414.1 GI:5941387
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 450)
AUTHORS Bujard,H., Gossen,M., Salfeld,J.G. and Voss,J.W.
TITLE Mice transgenic for a tetracycline-controlled transcriptional activator
JOURNAL Patent: US 5859310-A 5 12-JAN-1999;
FEATURES
  source
    Location/Qualifiers
      1..450
        /organism="unknown"
BASE COUNT      138 a      101 c      110 g      101 t
ORIGIN

Query Match      73.2%; Score 54.2; DB 6; Length 450;
Best Local Similarity 94.9%; Pred. No. 8.5e-11;
Matches 56; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAAACCGTC 60
      |||||||
Db 327 GCGGTGTACGGTGGGAGCCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAAGATCGCC 385

RESULT 6
AR029415
LOCUS AR029415
DEFINITION Sequence 6 from patent US 5859310.
ACCESSION AR029415
VERSION AR029415.1 GI:5941388
KEYWORDS
  
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:16 ; Search time 407.715 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: US-09-980-548-3
Perfect score: 74
Sequence: 1 tggcgtgtacgtgggagc.....accgtcaaacgcggaagct 74

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_be:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_em:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_be:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_or:*
- 21: em_ov:*
- 22: em_pat:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sts:*
- 27: em_un:*
- 28: em_vl:*
- 29: em_htg_hum:*
- 30: em_htg_inv:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vit:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	74	100.0	74	6	AX040911	AX040911 Sequence
2	54.2	73.2	432	6	I89986	I89986 Sequence 1
3	54.2	73.2	450	6	AR020199	AR020199 Sequence
4	54.2	73.2	450	6	AR020200	AR020200 Sequence
5	54.2	73.2	450	6	AR029414	AR029414 Sequence
6	54.2	73.2	450	6	AR029415	AR029415 Sequence
7	54.2	73.2	450	6	AR032151	AR032151 Sequence
8	54.2	73.2	450	6	AR032152	AR032152 Sequence
9	54.2	73.2	450	6	AR043816	AR043816 Sequence
10	54.2	73.2	450	6	AR043817	AR043817 Sequence
11	54.2	73.2	450	6	AR072128	AR072128 Sequence
12	54.2	73.2	450	6	AR072129	AR072129 Sequence
13	54.2	73.2	450	6	AR095982	AR095982 Sequence
14	54.2	73.2	450	6	AR095983	AR095983 Sequence
15	54.2	73.2	450	6	AR136642	AR136642 Sequence
16	54.2	73.2	450	6	AR136643	AR136643 Sequence
17	54.2	73.2	450	6	AR157177	AR157177 Sequence
18	54.2	73.2	450	6	AR157178	AR157178 Sequence
19	54.2	73.2	450	6	AR164126	AR164126 Sequence
20	54.2	73.2	450	6	AR164127	AR164127 Sequence
21	54.2	73.2	450	6	AX127249	AX127249 Sequence
22	54.2	73.2	450	6	AX127250	AX127250 Sequence
23	54.2	73.2	450	6	I15364	I15364 Sequence 12
24	54.2	73.2	450	6	I15365	I15365 Sequence 13
25	54.2	73.2	450	6	I32812	I32812 Sequence 8
26	54.2	73.2	450	6	I32813	I32813 Sequence 9
27	54.2	73.2	450	6	I56756	I56756 Sequence 5
28	54.2	73.2	450	6	I56757	I56757 Sequence 6
29	54.2	73.2	450	6	I59626	I59626 Sequence 8
30	54.2	73.2	450	6	I59627	I59627 Sequence 9
31	54.2	73.2	481	6	AR001287	AR001287 Sequence
32	54.2	73.2	481	6	AR116269	AR116269 Sequence
33	54.2	73.2	520	6	AR020198	AR020198 Sequence
34	54.2	73.2	520	6	AR032150	AR032150 Sequence
35	54.2	73.2	520	6	AR043815	AR043815 Sequence
36	54.2	73.2	520	6	AR072127	AR072127 Sequence
37	54.2	73.2	520	6	AR095981	AR095981 Sequence
38	54.2	73.2	520	6	AR136641	AR136641 Sequence
39	54.2	73.2	520	6	AR157176	AR157176 Sequence
40	54.2	73.2	520	6	AR164125	AR164125 Sequence
41	54.2	73.2	520	6	AX127248	AX127248 Sequence
42	54.2	73.2	520	6	I32811	I32811 Sequence 7
43	54.2	73.2	520	6	I59625	I59625 Sequence 7
44	54.2	73.2	569	6	AR020197	AR020197 Sequence
45	54.2	73.2	569	6	AR032149	AR032149 Sequence

ALIGNMENTS

RESULT 1	AX040911	AX040911	74 bp	DNA	linear	PAT 23-NOV-2000
LOCUS	Sequence 3 from Patent WO0065074.					
DEFINITION	AX040911					
ACCESSION	AX040911.1	GI:11340533				
VERSION						
KEYWORDS	unidentified.					
SOURCE	unidentified					
ORGANISM	unclassified					
REFERENCE	1 (bases 1 to 74)					
AUTHORS	Mueller.P., Geserick.C., Schroeder.K. and Hauser.H.					
TITLE	Promoter-transactivator system for inducible high-level mammalian gene expression with the option of cell growth control					
JOURNAL	Patent: WO 0065074-A 3 02-NOV-2000;					

Search completed: March 6, 2003, 00:22:21
Job time : 533.208 secs

Query Match 39.4%; Score 27.6; DB 14; Length 753;
Best Local Similarity 67.2%; Pred. No. 1.3e+02;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 12 GCGAATGGAAGTCAATCCCTCTCTCGGGAATGGAATCGAAATCGAAT 69
DB 452 GCGAGAGGAAGTGTGCTCAGAGCTCTGGAAGCAAGGAGACTGAAATCGAAT 509

RESULT 13
LOCUS AV096262 279 bp mRNA linear EST 22-NOV-1999
DEFINITION AV096262 Mus musculus C57BL/6J ES cell Mus musculus CDNA clone
ACCESSION AV096262
VERSION AV096262.1 GI:5243810
KEYWORDS house mouse.
SOURCE Mus musculus

REFERENCE
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Naitsumu, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE
JOURNAL RIKEN Mouse ESTs
COMMENT Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp

Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
source
Location/Qualifiers
1..279
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2410017M04"
/clone_lib="Mus musculus C57BL/6J ES cell"
/cell_type="ES cell"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACCAATCGAATGGAGCGCGGCGGGAATGGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 54 c 61 g 74 t
ORIGIN
Query Match 39.1%; Score 27.4; DB 9; Length 279;
Best Local Similarity 69.8%; Pred. No. 1.3e+02;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 12 GCGAATGGAAGTCAATCCCTCTCTCGGGAATGGAATCGAAAT 64

Db 5 GCGAATGGAACCTATACCTTCCATCCATTTACAGGAATGGTGACCAAT 57

RESULT 14
LOCUS AO664260 451 bp DNA linear GSS 23-JUN-1999
DEFINITION HS-596_B1_F11_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-1072 Col-21 Row-L, DNA sequence.
ACCESSION AO664260
VERSION AO664260.1 GI:5171951
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 451)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pletier de Jong (pletier@jeff.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1072 row: L column: 21
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 451.
Location/Qualifiers
1..451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate:1072 Col-21 Row=L"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 121 a 119 c 100 g 111 t
ORIGIN
Query Match 39.1%; Score 27.4; DB 17; Length 451;
Best Local Similarity 69.8%; Pred. No. 1.4e+02;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 2 ATCCCTTCGCGGAATCGAAATCGAAATCGAATCGAATCGAATGGA 54
DB 123 ATGCCTTCTCACTAAATGCGCAAGGAAAGTAGGATTCTCTCTTGGATTGGA 175

RESULT 15
LOCUS BI938198 747 bp mRNA linear EST 18-OCT-2001
DEFINITION de36c01.y1 wellcome CRC prn3 lip xenopus laevis cDNA clone IMAGE:3473665 5' similar to TR:014597 014597 NON-FUNCTIONAL FOLATE BINDING PROTEIN. ;, mRNA sequence.
ACCESSION BI938198
VERSION BI938198.1 GI:16252670

```

/Note=vector: pSPORI-P (Invitrogen technologies); Site-1
/Note: Site_2: MluI; mRNA obtained from wheat seeds of
cultivar Glenlea 5 days post-anthesis"
204 a lll c 150 q 193 t
BASE COUNT

```

```

/Note=vector: pSPORI-P (Invitrogen technologies); Site-1
/Note: Site_2: MluI; mRNA obtained from wheat seeds of
cultivar Glenlea 5 days post-anthesis"
204 a lll c 150 q 193 t
BASE COUNT

```

```

NOTI; Site_2: MluI; mRNA obtained from wheat seeds of
cultivar Glenlea 5 days post-anthesis"
BASE COUNT      204 a    111 c    150 g    193 t
ORIGIN
Query Match      39.4%; Score 27.6; DB 14; Length 658;
Best Local Similarity 67.2%; Pred. No. 1.3e+02;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 12 GCGAATGGAAACTGAAATCATCAGCTCTCGGGAAATCGAAATCGAATTCAGAT 69
    ||| ||||| ||| ||||| ||| ||| ||| ||||| |||||
Db 324 GCGAGGAGGAAGTGTGCTCAGAGCTCTCGAGCAAAAGGGAGACTGAAATTCAGAT 381
RESULT 12
BQ842141
LOCUS
DEFINITION
WHE2989_D07_H13Zs Wheat dormant embryo cDNA library Triticum
aestivum cDNA clone WHE2989_D07_H13, mRNA sequence.
ACCESSION
BQ842141
VERSION
BQ842141.1 GI:22211550
KEYWORDS
EST.
SOURCE
bread wheat.

```

REYNOLDS	EST.	SOURCE
		bread wheat.
		<i>Triticum aestivum</i>
		ORGANISM

REFERENCE
AUTHORS
TITLE

Anderson, O.D., Chao, S., Chin, A., Close, T.J., Doherty, L., Fenton, R.D., Lazo, G.R., Rausch, C.J., Walker-Simmons, M.K. and Wilson, C. The structure and function of the expressed portion of the wheat genomes - Dormant embryo cDNA library

1 (bases 1 to 753)
; Triticeae; Triticum.

zakaria; viridiplantae; Streptophyta; Embryophyta; tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae

CONTACT: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: andersn@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: SK primer

```

FEATURES
source
Location/Qualifiers
1. .753
/organism="Triticum aestivum"
/cultivar="Brevor"
/db xref="taxon:4565"

```

```

/cultivar="Brevor"
/db_xref="taxon:4565"
/clone="WHE2989_D07_H13"
/clone_lib="Wheat dormant embryo cDNA library"
/tissue_type="Seed embryo"
/dev_stage="Mature seed"

```

/tissue_type="seed embryo"
 /dev_stage="Mature seed"
 /lab_host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: ECOLI; Site_2: XhoI; Plants were grown to seed
 maturity under conditions favoring seed dormancy (L.
 Doherty at K. Walker-Simmons lab, Washington State
 University, Pullman, WA). Embryos were cut from mature
 dormant seed (Doherty). Total RNA was prepared from these
 embryos, polyA was purified, a cDNA library was made, and
 the cDNA clones were in vivo excised to give pBluescript
 phagemids in the T7 Clonase lab at the University of

phagemids in the TU Close lab at the University of California, Riverside (Chin, Fenton). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

[illegible]

QY	54	TCACA	68
DB	104	GCACA	100

RESULT 8

RESULT 8
T97467/C


```

/db_xref="taxon:10090"
/clone_image="IMAGE:5645803"
/clone_lib="Melton Mouse E16.5 Pancreas Library 2 M1682"
/sex="Both"
/tissue_type="Total pancreas"
/dev_stage="Embryonic day 16.5"
/lab_host="TOP10"
/note="Organ: Pancreas; Vector: pBluescript II SK; Site:1: NotI; Site:2: SalI; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming.Size-selected by column fractionation; average insert size 1.06kb. Primary library, unamplified."

```

BASE COUNT	152 a	128 c	99 g	161 t
ORIGIN				

Query Match	42.9%	Score 30;	DB 12;	Length 540;
Best Local Similarity	64.3%;	pred. No. 23;		
Matches	45;	Conservative	0;	Mismatches 25;
				Indels 0;
				Gaps 0;

```

Query Match      45.1%; Score 31.6; DB 14; Length 506;
Best Local Similarity 74.1%; Pred. No. 7.3;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

RESULT 3
BI897688/c
LOCUS
BI897688 592 bp mRNA linear EST 15-OCT-2001
DEFINITION
f04e10.x1 Zebrafish Research Genetics C32 fin Danio rerio cDNA
clone 4463466 3' similar to TR:014597 014597 NON-FUNCTIONAL POLYATE
BINDING PROTEIN. ; mRNA sequence.

BI897688
BI897688.1 GI:16140824

ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 592)
REFERENCE

REFERENCE
AUTHORS
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
1 (bases 1 to 532).
S. Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
Cyprine, J. (1980).
1 (bases 1 to 532).

TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu

e-mail: zbrar@shewatson.wustl.edu
 cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by: Steve
 Johnson. DNA Sequencing by: Washington University Genome Sequencing
 Center. Clone distribution: IMAGE consortium email:
info@image.llnl.gov
 Possible reversed clone: polyT not found
 Seq primer: T7 from Gibco
 High quality sequence stop: 469.

```

/seq_xref="taxon:7953"
/clone="4469466"
/clone_lib="Zebrafish Research Genetics C32 fin"
/tissue_type="Fin"
/lab_host="GeneHogs/HS996, a phage-resistant isolate of

```

```

/clone_lib="Zebrafish Research Genetics C32 lin"
/tissue_type="Fin"
/lab_host="GeneHoops (HS996, a phage-resistant isolate of

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 527.874 Seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-2
Perfect score: 70
Sequence: 1 gatcccttcggaatgg.....tggaactgaaatcagatc 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Jeached: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_oth:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.6	45.1	506	14	BQ321577
2	30	42.9	540	12	BG142899
3	29.8	42.6	592	13	BI897688
4	29.6	42.3	517	17	AZ471549
5	28.2	40.3	391	13	BI936879
6	28.2	40.3	391	17	AZ913831

c	7	27.8	39.7	391	14	R94003
c	9	27.6	39.4	434	14	T97467
	10	27.6	39.4	502	17	AG630605
	11	27.6	39.4	656	13	BJ255893
	12	27.6	39.4	658	14	BQ238956
	13	27.4	39.1	753	14	BQ842141
	14	27.4	39.1	279	9	AV096362
	15	27.4	39.1	451	17	AQ664260
c	16	27.2	38.9	747	13	BI938198
	17	27.2	38.9	688	10	AV681534
c	18	27	38.6	1037	12	BF979744
	19	27	38.6	419	14	N59640
	20	27	38.6	695	17	AZ812128
	21	26.8	38.3	719	13	BI855882
	22	26.8	38.3	524	13	BJ032573
c	23	26.8	38.3	533	17	AQ623869
c	24	26.8	38.3	569	17	AQ695555
c	25	26.8	38.3	605	10	AV825378
c	26	26.8	38.3	655	17	BH674985
c	27	26.8	38.3	682	17	BH680864
c	28	26.8	38.3	710	12	BF119479
c	29	26.6	38.0	164	17	AZ745116
c	30	26.6	38.0	243	10	AV355881
c	31	26.6	38.0	331	9	AA728941
c	32	26.6	38.0	341	12	BG508795
c	33	26.6	38.0	473	17	AZ159253
	34	26.6	38.0	550	17	AQ630596
	35	26.6	38.0	575	13	BM383892
c	36	26.6	38.0	1059	13	BM467331
c	37	26.4	37.7	418	17	AQ210467
c	38	26.4	37.7	508	13	BI940410
c	39	26.4	37.7	726	9	AI217912
c	40	26.2	37.4	225	10	BB337449
c	41	26.2	37.4	481	14	T56182
c	42	26.2	37.4	534	12	BG447023
c	43	26.2	37.4	603	13	BI503827
c	44	26.2	37.4	668	13	BJ071703
c	45	26	37.1	184	10	AW021778

ALIGNMENTS

RESULT 1
BQ321577
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

BQ321577
RC1-CT0249-130900-212-b06 CT0249 Homo sapiens cDNA, mRNA sequence.
506 bp
linear
EST 17-MAY-2002

EST.
human.
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 506)
Dias Neto,B., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Cosca,F.F., Goldman,G.H., Carvalho,A.F., Matsumura,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Thu Mar 6 09:56:00 2003

us-09-980-548-2.rng

Page 12

Search completed: March 5, 2003, 22:49:56
Job time : 73.3092 secs

PS Claim 25; SEQ ID No 4962; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

CC (SENPs). The present sequence is one such probe. The SENPs are derived

CC from human Hela cells. The SENPs can be used to produce a single exon

CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene

CC expression, the probes are therefore useful in grading and/or staging

CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

XX

XX Query Match 36.6%; Score 25.6; DB 22; Length 415;

XX Best Local Similarity 66.1%; Pred. No. 23;

XX Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

XX

QY 15 AAATGGAACCTGAAATCAGATCCCTCTCGGGAATGGAACCTGAAATCAGATC 70

DB 201 AAATGACATAAAATTCATCCATCCATGACCTAAATGGAATGGAATCAGATC 256

IIII II III III III III I III III III III III III

RESULT 14

AAI36372

ID AAI36372 standard; DNA; 415 BP.

XX

AC AAI36372;

XX

DT 17-OCT-2001 (first entry)

XX

DE Probe #5058 used to measure gene expression in human placenta sample.

XX

DE Probe; microarray; human; placenta; antenatal diagnosis;

XX

KW genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200157272-A2.

XX

PD 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WO-US00663.

XX

XX 04-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 03-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234687.

XX

XX 27-SEP-2000; 2000US-0236359.

XX

XX 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-488897/53.

XX

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

XX

XX Claim 25; SEQ ID No 5058; 654pp; English.

XX

XX The present invention relates to single exon nucleic acid probes (SENPs).

XX The present sequence is one such probe. The probes are useful for

XX producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful

XX for antenatal diagnosis of human genetic disorders.

XX

XX Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

XX

XX Query Match 36.6%; Score 25.6; DB 22; Length 415;

Best Local Similarity 66.1%; Pred. No. 23;

Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 15 AAATGGAACCTGAAATCAGATCCCTCTCGGGAATGGAACCTGAAATCAGATC 70

DB 201 AAATGACATAAAATTCATCCATCCATGACCTAAATGGAATGGAATCAGATC 256

IIII II III III III III I III III III III III III

RESULT 15

AAI04795

ID AAI04795 standard; DNA; 415 BP.

XX

AC AAI04795;

XX

DT 09-OCT-2001 (first entry)

XX

DE Probe #4786 used to measure gene expression in human breast sample.

XX

DE Probe; human; breast disease; breast cancer; development disorder; ss;

XX

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX

OS Homo sapiens.

XX

PN WO200157270-A2.

XX

PD 09-AUG-2001.

XX

XX 29-JAN-2001; 2001WO-US00661.

XX

XX 04-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 03-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234687.

XX

XX 27-SEP-2000; 2000US-0236359.

XX

XX 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-476286/51.

XX

XX Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

XX

XX Claim 25; SEQ ID No 4786; 322pp; English.

XX

XX The present invention relates to novel single exon nucleic acid probes.

XX The present sequence is one such probe. The probes are useful for

XX measuring human gene expression in a human breast sample, where the probe

XX hybridises at high stringency to a nucleic acid expressed in the human

XX breast. The probes are useful for predicting, diagnosing, grading,

XX staging, monitoring and prognosing diseases of the human breast,

XX particularly those diseases with polygenic aetiology. The diseases

XX include: breast cancer, disorders of development, inflammatory diseases

XX of the breast, fibrocystic changes, proliferative breast disease and

XX non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 415 BP; 136 A; 91 C; 80 G; 108 T; 0 other;

XX

XX Query Match 36.6%; Score 25.6; DB 22; Length 415;

XX Best Local Similarity 66.1%; Pred. No. 23;

XX Matches 37; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

XX

QY 15 AAATGGAACCTGAAATCAGATCCCTCTCGGGAATGGAACCTGAAATCAGATC 70

DB 201 AAATGACATAAAATTCATCCATCCATGACCTAAATGGAATGGAATCAGATC 256

IIII II III III III III I III III III III III III

RESULT 10
ABA26396

PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 28-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225288.
 PR 14-AUG-2000; 2000US-0225289.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-502630/55.
 XX
 PT Polynucleotides encoding digestive system antigens, useful for
 PT diagnosing, treating, preventing and/or prognosing disorders of the
 PT digestive system, particularly cancer and cancer metastases -
 XX
 PS Disclosure; SEQ ID NO 4997; 986pp; English.
 XX
 CC The present invention provides the protein and coding sequences of a

CC (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC colonic inflammation, diarrhoea and dysentery, malabsorption syndromes,
CC (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC The polynucleotide sequences of the invention can also be used in gene
CC therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC novel human colon associated polypeptides of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 551 BP; 187 A; 101 C; 116 G; 147 T; 0 other;

Query Match 38.6%; Score 27; DB 22; Length 551;
Best Local Similarity 66.1%; Pred. No. 8.1;
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 11 CGGGAATGGAACAACTGAAATCAGATCCCTTCGCGGAATGGAACCTGAAATCAGAT 69
DB 193 CTGCTACTGAAACAAACAAATAGACCCCTGTGTGAGAAAGGAAATATGAAATCAGAT 251

RESULT 4

40004

AAS40004 standard; DNA; 551 BP.

AAS40004;

DT 17-DEC-2001 (first entry)

DE Genomic sequence #423 encoding human colon associated polypeptide.

Human; colon cancer; congenital abnormality; infection; colitis;
KW inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW cytosolic; ds.

OS Homo sapiens.

PN WO200155302-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01240.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0215647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:12 ; Search time 68.3092 Seconds
(without alignments)
2307.741 Million cell updates/sec

Title: US-09-980-548-2

Perfect score: 70

Sequence: 1 gatcccttcgggaatgg.....tggaactgaaatcagatc 70

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

..searched: 2185239 seqs, 1125995159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	70	100.0	70	21	Firefly IRF-1 bind
2	27.6	39.4	2286	23	DNA encoding novel
3	27	38.6	551	22	Genomic sequence #
4	27	38.6	551	22	Genomic sequence #
5	27	38.6	551	22	Human digestive sy
6	27	38.6	551	22	Human digestive sy
c 7	26.2	37.4	481	24	Human ovarian canc
8	25.6	36.6	415	22	Human breast cell
9	25.6	36.6	415	22	Human foetal liver

10	25.6	36.6	415	22	ABA26396
11	25.6	36.6	415	22	AAK04899
12	25.6	36.6	415	22	AAK30423
13	25.6	36.6	415	22	AAI15029
14	25.6	36.6	415	22	AAI16372
15	25.6	36.6	415	22	AAI04795
16	25.6	36.6	415	22	ABS05034
17	25.6	36.6	8981	21	AAAC69138
c 18	25.6	36.6	183999	22	AAF92831
c 19	25.6	36.6	715	20	AAI37514
20	25.4	36.3	90104	23	ABL12402
21	25.4	36.3	684707	24	ABQ67196
22	25.4	36.3	3011208	24	ABQ69245
c 23	25.2	36.0	348	22	AAQ20335
c 24	25.2	36.0	1230025	20	AAI91990
25	25	35.7	976	23	ABV04483
26	24.8	35.4	310	20	AAV89187
c 27	24.8	35.4	685	21	AAI33345
28	24.8	35.4	705	22	AAF22680
29	24.8	35.4	887	22	AAF22682
30	24.8	35.4	996	21	AAI05586
31	24.8	35.4	1789	20	AAI40059
32	24.8	35.4	2353	19	AAV40542
33	24.8	35.4	2566	22	AAF59641
c 34	24.8	35.4	7568	23	ABL14768
c 35	24.8	35.4	11887	19	AAV52279
c 36	24.8	35.4	49312	21	AAH51594
c 37	24.8	35.4	140036	24	AAI98600
c 38	24.6	35.1	1510	22	AAH99084
c 39	24.6	35.1	3884	22	AAH98627
c 40	24.6	35.1	5397	23	AAI56600
41	24.6	35.1	9909	19	AAV52145
42	24.6	35.1	2365589	24	ABA90521
43	24.4	34.9	244	22	AAI32933
44	24.4	34.9	244	22	AAH82491
c 45	24.4	34.9	1555	21	AAI41543

ALIGNMENTS

RESULT 1
AAC64592

ID AAC64592 standard; DNA; 70 BP.

AC AAC64592;

XX AAC64592;

XX AAC64592;

DT 15-FEB-2001 (first entry)

DE Firefly IRF-1 binding site nucleotide sequence SEQ ID NO:2.

XX Myeloproliferative sarcoma virus; MPSV; MPSV-E; enhancer; CMV;
XX cytomegalovirus; firefly; IRF-1 binding site; minimal promoter;
XX interferon regulatory factor 1 binding site; transcription regulator;
XX promoter transactivator system; ds.

OS Photinus pyralis.

XX Photinus pyralis.

XX Photinus pyralis.

PN EPI046710-A1.

XX EPI046710-A1.

PD 25-OCT-2000.

XX 25-OCT-2000.

XX 23-APR-1999; 99EP-0108068.

XX 23-APR-1999; 99EP-0108068.

XX (GBFB) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.

PI Mueller PP, Geserick C, Schroeder K, Hauser H;

XX WPI; 2000-648930/63.

PT Promoter-transactivator system, useful for inducing high level

```
misc_feature 5440. .5449 /note="match to EST BF354907 (NID:g11313981)"
repeat_region 7059. .7366 /rpt_family="Alu"
repeat_region 7340. .7386 /rpt_family="(A)n"
repeat_region 7392. .7550 /rpt_family="L1"
misc_feature 7719. .8230 /note="match to EST AI459058 (NID:g4311637) tk03e05.x1"
repeat_region 8265. .8388 /rpt_family="L2"
repeat_region 9415. .9466 /rpt_family="(TA)n"
repeat_region 10207. .10258 /rpt_family="(TTTA)n"
repeat_region 10232. .10370 /rpt_family="Alu"
repeat_region 11214. .11365 /rpt_family="MIR"
repeat_region 11400. .11620 /rpt_family="MIR"
repeat_region 11641. .11932 /rpt_family="Alu"
repeat_region 12012. .12105 /rpt_family="MIR"
repeat_region 12631. .12754 /rpt_family="MIR"
repeat_region 13218. .13336 /rpt_family="MIR"
repeat_region 13328. .13569 /rpt_family="L2"
repeat_region 13560. .13648 /rpt_family="MIR"
repeat_region 13688. .13821 /rpt_family="MIR"
repeat_region 13907. .14441 /rpt_family="L2"
repeat_region 14542. .14580 /rpt_family="L2"
misc_feature 15070. .15475 /note="match to EST AI032114 (NID:g3250326) os76a07.s1"
repeat_region 15239. .15525 /rpt_family="Alu"
misc_feature 15610. .15611 /note="match to EST AI032114 (NID:g3250326) os76a07.s1"
repeat_region 16114. .16234 /rpt_family="L1"
repeat_region 16517. .16784 /rpt_family="L1"
repeat_region 16787. .16863 /rpt_family="MER2_type"
repeat_region 16844. .16907 /rpt_family="AT_rich"
repeat_region 16886. .17188 /rpt_family="Alu"
repeat_region 17246. .17352 /rpt_family="MER2_type"
repeat_region 17353. .17643 /rpt_family="L1"
repeat_region 17657. .18139 /rpt_family="L1"
repeat_region 18140. .18990 /rpt_family="L1"
repeat_region 18617. .18665 /rpt_family="AT_rich"
repeat_region 19048. .19330 /rpt_family="L1"
repeat_region 19333. .20537 /rpt_family="MER2_type"
repeat_region 20539. .20717 /rpt_family="Alu"
repeat_region 20718. .21877
```

```
repeat_region /rpt_family="MER2_type"
21880. .22318 /rpt_family="L1"
repeat_region 22315. .22543 /rpt_family="L1"
repeat_region 22572. .22606 /rpt_family="Mariner"
repeat_region 22616. .22975 /rpt_family="L1"
repeat_region 23377. .23609 /rpt_family="MIR"
repeat_region 24093. .24207 /rpt_family="L1"
repeat_region 24445. .24741 /rpt_family="Alu"
repeat_region 24719. .24741 /rpt_family="(A)n"
misc_feature 25224. .25290 /note="match to EST AL561641 (NID:g12909272)"
misc_feature 25225. .25641 /note="match to EST BG192864 (NID:g13714551)"
misc_feature 25230. .25290 /note="match to EST BE561346 (NID:g9805053)"
misc_feature 25230. .25290

Query Match 41.48; Score 29; DB 9; Length 165373;
Best Local Similarity 63.8%; Pred.No. 41;
Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GATCCCTTCTCGGGAATGGAATCAGATCCCTTCTCGGGAATGGAATGGAATGGA 60
DB 113685 GACCCCTATCTCTTAAAAAACAAGAAAAATATCACTTCTCGGGAACAGAACTTAA 113744

QY 61 AAATCAGAT 69
DB 113745 AGACACAGAT 113753

Search completed: March 5, 2003, 23:34:38
Job time : 766.676 secs
```



```

misc_feature      /note="assembly_fragment"
2879. 6107
misc_feature      /note="assembly_fragment"
6208. 7623
misc_feature      /note="assembly_fragment"
clone_end:SP6
vector_side:right
7724. 11841
misc_feature      /note="assembly_fragment"
clone_end:T7
vector_side:right
11942. 18456
misc_feature      /note="assembly_fragment"
18557. 27382
misc_feature      /note="assembly_fragment"
27483. 37124
misc_feature      /note="assembly_fragment"
37225. 47276
misc_feature      /note="assembly_fragment"
47377. 68713
misc_feature      /note="assembly_fragment"
68814. 89140
misc_feature      /note="assembly_fragment"
89241. 116360
misc_feature      /note="assembly_fragment"
116461. 141816
misc_feature      /note="assembly_fragment"
141917. 176775
misc_feature      /note="assembly_fragment"
42683 a 45397 c 44545 g 42948 t 1202 others
ORIGIN
Query Match      41.78; Score 29.2; DB 2; Length 176775;
Best Local Similarity 74.08; Pred. No. 35;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

Qy 13 GGAATGGAATGAAATCAGATCCCTCTCGGGAATGGAATGAACTGAAA 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 104829 GGAATGGAGACAGCAAGTCAGACTCTCCCTGTAAGGGAAGAGAAA 104780

```

```

RESULT 15
AC067957          AC067957      165373 bp      DNA      linear      PRI 09-JAN-2002
LOCUS            Homo sapiens BAC clone RP11-82K13 from 2, complete sequence.
DEFINITION       AC067957
ACCESSION        AC067957
VERSION          AC067957.7 GI:15638744
KEYWORDS         HTG.
SOURCE           Homo sapiens.
ORGANISM         Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Isak.A., Kozlowicz.A., Dixon.R. and Doeberer,A.
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
95063792
9847074
PUBMED
2 (bases 1 to 165373)
Isak.A., Kozlowicz.A., Dixon.R. and Doeberer,A.
The sequence of Homo sapiens BAC clone RP11-82K13
Unpublished (2001)
3 (bases 1 to 165373)
Waterston,R.H.
Direct Submission
Submitted (27-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 165373)
Waterston,R.H.
Direct Submission
Submitted (18-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

MO 63108, USA
5 (bases 1 to 165373)
Waterston,R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 18, 2001 this sequence version replaced gi:14670148.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_NH0082K13
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frenken,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-111K7; the clone sequenced to the right is RP11-89K21, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-82K13; actual end is at base position 29707 of RP11-89K21.

Data from AC012455, AC067791, and AC062030 was used to finish this clone, AC067957.

Polymorphisms exist between AC067957 and AC067791.

```

FEATURES
source
    Location/Qualifiers
        1..165373
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2"
            /clone="RP11-82K13"
            /clone_lib="RPCI-11"
            3097..3427
            /note="match to EST H64341 (NID:g1023081) yr71a12.rl"
misc_feature
    repeat_region
        4326..4440
            /rpt_family="MIR"
repeat_region
    4486..4799
        /rpt_family="Alu"
repeat_region
    4770..4820
        /rpt_family="(A)n"
misc_feature
    5390..5549
        /note="match to EST AA485569 (NID:g2214788) ab09g06.rl"

```

<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-149P14. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP11-149P14 is at 1 in this sequence. The true left end of clone RP11-149P14 is at 108122 in this sequence. The true right end of clone RP11-192B7 is at 58622 in this sequence.

```

FEATURES             Location/Qualifiers
     source            1..110121
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="1"
                        /clone="RP11-149P14"
                        /clone_lib="RPCI-11.1"
     misc_feature      6650..6686
                        /note="Sequence from overlapping clone RP11-192B7
                        (AL592314)..Assembly confirmed by restriction digest."
     misc_feature      21058..21123
                        /note="Sequence from overlapping clone RP11-192B7
                        (AL592314)..Assembly confirmed by restriction digest."
     misc_feature      50716..50718
                        /note="Sequence from overlapping clone RP11-192B7
                        (AL592314)..Assembly confirmed by restriction digest."
     misc_feature      54763..54871
                        /note="Sequence from overlapping clone RP11-192B7
                        (AL592314)..Assembly confirmed by restriction digest."
     misc_feature      86404..86414
                        /note="Sequence from uni-directional dGTP big dye
                        terminator reads only."
     misc_feature      86780..86903
                        /note="Single clone region. Reads generated from a
                        transposon library derived from a single pUC clone.
                        Restriction digest data confirm the assembly."
BASE COUNT    25978 a 28405 c 27853 g 27885 t
ORIGIN
Query Match          41.7%; Score 29.2; DB 9; Length 110121;
Best Local Similarity 74.0%; Pred. No. 35;
Matches 37; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 13 GGAATGGAACCTGAATCAGACTCTCTCGGGAATCGAACTGAAA 62
||||| ||||| || ||||| ||||| ||||| ||||| |||||
Db 26583 GGAATTGGAGACAGCAAGTCAGAACTCTCCCTGTAAGGGAAGAGAAA 26534

RESULT 14
AC016978/c          AC016978          176775 bp      DNA      linear      HTG 09-SEP-2000
LOCUS              Homo sapiens clone RP11-5N7, WORKING DRAFT SEQUENCE, 13 unordered
DEFINITION          pieces.
ACCESSION           AC016978
VERSION             AC016978.3 GI:10045327
KEYWORDS             HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE              Homo sapiens.
ORGANISM            Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 176775)
AUTHORS             Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
                    Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
                    Boguslavskiy,L., Boukhgalter,B., Brown,A., Castile,A., Colangelo,M.,
                    Collins,S., Collymore,A., Cooke,P., DeArelano,K., Dewar,K.,
                    Domino,M., Doyle,M., Fenestor,J., Ferreira,P., FitzHugh,W.,
                    Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,
                    Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,
                    Karatsas,A., Klein,J., Landers,T., Lehoczy,J., Lieu,C., Locke,K.,
                    MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
                    Mcdonald,P., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,
                    O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,
                    Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,
                    Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
                    Tirrell,A., Vassiliev,H., Viel,R., VO.A., Wu,X., Wyman,D., Ye.W.J.,
                    Zimmer,A. and Zody,M.
                    Direct Submission
                    Submitted (09-DEC-1999) Whitehead Institute/MIT Center for Genome
                    Research, 320 Charles Street, Cambridge, MA 02141, USA
                    On Sep 9, 2000 this sequence version replaced gi:6648261.
                    All repeats were identified using RepeatMasker:
                    Smit, A.F.A. & Green, P. (1996-1997)
                    http://ftp.genome.washington.edu/RM/RepeatMasker.html
                    ----- Genome Center
                    Center: Whitehead Institute/ MIT Center for Genome Research
                    Center code: WIBR
                    Web site: http://www-seq.wi.mit.edu
                    Contact: sequence_submissions@genome.wi.mit.edu
                    ----- Project Information
                    Center project name: L2875
                    Center clone name: 5_N7
                    ----- Summary Statistics
                    Sequencing vector: M13; M77815; 100% of reads
                    Chemistry: Dye-terminator Big Dye; 100% of reads
                    Assembly program: Phrap; version 0.960731
                    Consensus quality: 166773 bases at least Q40
                    Consensus quality: 172242 bases at least Q30
                    Consensus quality: 174181 bases at least Q20
                    Insert size: 85000; agarose-fp
                    Insert size: 175575; sum-of-contigs
                    Quality coverage: 8.8 in Q20 bases; agarose-fp
                    Quality coverage: 4.3 in Q20 bases; sum-of-contigs
                    -----
                    * NOTE: This is a 'working draft' sequence. It currently
                    * consists of 13 contigs. The true order of the pieces
                    * is not known and their order in this sequence record is
                    * arbitrary. Gaps between the contigs are represented as
                    * runs of N, but the exact sizes of the gaps are unknown.
                    * This record will be updated with the finished sequence
                    * as soon as it is available and the accession number will
                    * be preserved.
                    *
                    * 1 2778: contig of 2778 bp in length
                    * 2779 2878: gap of 100 bp
                    * 2879 6107: contig of 3229 bp in length
                    * 6108 6207: gap of 100 bp
                    * 6208 7623: contig of 1416 bp in length
                    * 7624 7723: gap of 100 bp
                    * 7724 11841: contig of 4118 bp in length
                    * 11842 11941: gap of 100 bp
                    * 11942 18456: contig of 6515 bp in length
                    * 18457 18556: gap of 100 bp
                    * 18557 27382: contig of 8826 bp in length
                    * 27383 27482: gap of 100 bp
                    * 27483 37124: contig of 9642 bp in length
                    * 37125 37224: gap of 100 bp
                    * 37225 47276: contig of 10052 bp in length
                    * 47277 47376: gap of 100 bp
                    * 47377 66713: contig of 19337 bp in length;
                    * 66714 66813: gap of 100 bp
                    * 66814 89140: contig of 22327 bp in length
                    * 89141 89240: gap of 100 bp
                    * 89241 116360: contig of 27120 bp in length
                    * 116361 116460: gap of 100 bp
                    * 116461 141816: contig of 25356 bp in length
                    * 141817 141916: gap of 100 bp
                    * 141917 176775: contig of 34859 bp in length.
                    Location/Qualifiers
                    1..176775
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="RP11-5N7"
                    /clone_lib="RPCI-11 Human Male BAC"
                    misc_feature 1..2778
FEATURES             source

```

Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., VO.A., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-DEC-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 9, 2000 this sequence version replaced gi:6648261.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2875
Center clone name: 5_N7
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166773 bases at least Q40
Consensus quality: 172242 bases at least Q30
Consensus quality: 174181 bases at least Q20
Insert size: 85000; agarose-fp
Insert size: 175575; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 4.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 13 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 2778: contig of 2778 bp in length

* 2779 2878: gap of 100 bp

* 2879 6107: contig of 3229 bp in length

* 6108 6207: gap of 100 bp

* 6208 7623: contig of 1416 bp in length

* 7624 7723: gap of 100 bp

* 7724 11841: contig of 4118 bp in length

* 11842 11941: gap of 100 bp

* 11942 18456: contig of 6515 bp in length

* 18457 18556: gap of 100 bp

* 18557 27382: contig of 8826 bp in length

* 27383 27482: gap of 100 bp

* 27483 37124: contig of 9642 bp in length

* 37125 37224: gap of 100 bp

* 37225 47276: contig of 10052 bp in length

* 47277 47376: gap of 100 bp

* 47377 66713: contig of 19337 bp in length;

* 66714 66813: gap of 100 bp

* 66814 89140: contig of 22327 bp in length

* 89141 89240: gap of 100 bp

* 89241 116360: contig of 27120 bp in length

* 116361 116460: gap of 100 bp

* 116461 141816: contig of 25356 bp in length

* 141817 141916: gap of 100 bp

* 141917 176775: contig of 34859 bp in length.

Location/Qualifiers

1..176775

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="RP11-5N7"

/clone_lib="RPCI-11 Human Male BAC"

misc_feature 1..2778

* 25929 26799: contig of 871 bp in length
 * 26800 26899: gap of 100 bp
 * 26900 27808: contig of 909 bp in length
 * 27809 27908: gap of 100 bp
 * 27909 28812: contig of 904 bp in length
 * 28813 28912: gap of 100 bp
 * 28913 29813: contig of 901 bp in length
 * 29814 29913: gap of 100 bp
 * 29914 30807: contig of 894 bp in length
 * 30808 30907: gap of 100 bp
 * 30908 31790: contig of 883 bp in length
 * 31791 31890: gap of 100 bp
 * 31891 32777: contig of 887 bp in length
 * 32778 32877: gap of 100 bp
 * 32878 33783: contig of 912 bp in length
 * 33790 33889: gap of 100 bp
 * 33890 34792: contig of 903 bp in length
 * 34793 34892: gap of 100 bp
 * 34893 35795: contig of 903 bp in length
 * 35796 35895: gap of 100 bp
 * 35896 36798: contig of 903 bp in length
 * 36799 36898: gap of 100 bp
 * 36899 37791: contig of 893 bp in length
 * 37792 37891: gap of 100 bp
 * 37892 38732: contig of 841 bp in length
 * 38733 38832: gap of 100 bp
 * 38833 39745: contig of 914 bp in length
 * 39747 39846: gap of 100 bp
 * 39847 40745: contig of 899 bp in length
 * 40746 40845: gap of 100 bp
 * 40846 41743: contig of 898 bp in length
 * 41744 41843: gap of 100 bp
 * 41844 42723: contig of 880 bp in length
 * 42724 42823: gap of 100 bp
 * 42824 43687: contig of 864 bp in length
 * 43688 43787: gap of 100 bp
 * 43788 44683: contig of 896 bp in length
 * 44684 44783: gap of 100 bp
 * 44784 45688: contig of 902 bp in length
 * 45689 45785: gap of 100 bp
 * 45786 46689: contig of 904 bp in length
 * 46690 46789: gap of 100 bp
 * 46790 47690: contig of 901 bp in length
 * 47691 47790: gap of 100 bp
 * 47791 48690: contig of 900 bp in length
 * 48691 48790: gap of 100 bp
 * 48791 49709: contig of 919 bp in length
 * 49710 49809: gap of 100 bp
 * 49810 50684: contig of 875 bp in length
 * 50685 50784: gap of 100 bp
 * 50785 51684: contig of 900 bp in length
 * 51685 51784: gap of 100 bp
 * 51785 52685: contig of 901 bp in length
 * 52686 52785: gap of 100 bp
 * 52786 53691: contig of 906 bp in length
 * 53692 53791: gap of 100 bp
 * 53792 54688: contig of 897 bp in length
 * 54689 54788: gap of 100 bp
 * 54789 55703: contig of 917 bp in length
 * 55706 55805: gap of 100 bp
 * 55806 56714: contig of 909 bp in length
 * 56715 56814: gap of 100 bp
 * 56815 57704: contig of 890 bp in length
 * 57705 57804: gap of 100 bp
 * 57805 58703: contig of 898 bp in length
 * 58703 58802: gap of 100 bp
 * 58803 59683: contig of 881 bp in length
 * 59684 59783: gap of 100 bp
 * 59784 60662: contig of 879 bp in length
 * 60663 60762: gap of 100 bp
 * 60763 61642: contig of 880 bp in length
 * 61643 61742: gap of 100 bp
 * 61743 62617: contig of 875 bp in length

* 62618 62717: gap of 100 bp
 * 62718 63623: contig of 906 bp in length
 * 63624 63723: gap of 100 bp
 * 63724 64617: contig of 894 bp in length
 * 64618 64717: gap of 100 bp
 * 64718 65635: contig of 918 bp in length
 * 65636 65735: gap of 100 bp
 * 65736 66612: contig of 877 bp in length
 * 66613 66712: gap of 100 bp
 * 66713 67591: contig of 879 bp in length
 * 67592 67691: gap of 100 bp
 * 67692 68598: contig of 907 bp in length
 * 68599 68698: gap of 100 bp
 * 68699 69594: contig of 896 bp in length
 * 69595 69694: gap of 100 bp
 * 69695 70501: contig of 807 bp in length
 * 70502 70601: gap of 100 bp
 * 70602 71452: contig of 851 bp in length
 * 71453 71552: gap of 100 bp
 * 71553 72370: contig of 818 bp in length

Query Match 42.3%; Score 29.6; DB 2; Length 90747;
 Best Local Similarity 88.9%; Pred. No. 26;
 Matches 32; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 CCCTTCTCGGGAATGGAACCTGAAATCAGATCCC 39
 |||||
 Db 1466 CCCTTCTCGGGAATGGAACCTGAAATCTACTTCC 1431

RESULT 13
 AL590989/c
 LOCUS AL590989 Human DNA sequence from clone RP11-149P14 on chromosome 1, complete
 DEFINITION sequence.
 ACCESSION AL590989
 VERSION AL590989.15 GI:16973092
 KEYWORDS HTG
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 110121)
 AUTHORS Howden, P.
 TITLE Direct Submission
 JOURNAL Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 COMMENT On Nov 16, 2001 this sequence version replaced gi:16508264.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; SW; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chrl>
 RP11-149P14 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5392
Center clone name: 127_H_5

FEATURES

URES	source	Location/Qualifiers
		1. .160883
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="8"
		/map="8"
		/clone="RP11-127H5"
		/clone_lib="RPC1-11 Human Male BAC"
		2879. .3178
repeat_region		/rpt_family="AluSc"
repeat_region		3814. .3842
repeat_region		/rpt_family="AT_rich"
repeat_region		complement(4030. .4462)
repeat_region		/rpt_family="L2"
repeat_region		4486. .4540
repeat_region		/rpt_family="(TTCA)n"
repeat_region		complement(4550. .4830)
repeat_region		/rpt_family="L1MC1"
repeat_region		4831. .4920
repeat_region		/rpt_family="(CTATA)n"
repeat_region		complement(4927. .5161)
repeat_region		/rpt_family="L1MC1"
repeat_region		complement(6250. .6361)
repeat_region		/rpt_family="MIR"
repeat_region		6382. .6557
repeat_region		/rpt_family="MER63A"
repeat_region		complement(6558. .6698)
repeat_region		/rpt_family="MIR"
repeat_region		complement(6712. .6865)
repeat_region		/rpt_family="MIR3"
repeat_region		6901. .6927
repeat_region		/rpt_family="(T)n"
repeat_region		complement(8668. .9158)
repeat_region		/rpt_family="LTR33"
repeat_region		complement(9223. .9528)
repeat_region		/rpt_family="AluJb"
repeat_region		9962. .10210
repeat_region		/rpt_family="AluSk"
repeat_region		12393. .12953
repeat_region		/rpt_family="L2"
repeat_region		complement(13061. .13176)
repeat_region		/rpt_family="FLAM_A"
repeat_region		complement(13713. .13976)
repeat_region		/rpt_family="LTR16A1"
repeat_region		14722. .14907
repeat_region		/rpt_family="MIR"
repeat_region		14923. .15077
repeat_region		/rpt_family="MIR"
repeat_region		15547. .15742
repeat_region		/rpt_family="MIR"
repeat_region		complement(15987. .16070)
repeat_region		/rpt_family="L2"
repeat_region		complement(16162. .16216)
repeat_region		/rpt_family="MER103"
repeat_region		16855. .16901
repeat_region		/rpt_family="L2"
repeat_region		17188. .17240
repeat_region		/rpt_family="(TG)n"
repeat_region		complement(17279. .17346)
repeat_region		/rpt_family="L2"
repeat_region		complement(17721. .18015)
repeat_region		/rpt_family="AluY"
repeat_region		complement(18026. .18581)
repeat_region		/rpt_family="L1ME1"
repeat_region		complement(18749. .18863)

repeat_region	/rpt_family="FLAM_A" 18961. 19194
repeat_region	/rpt_family="MIR" complement(19347. 19420)
repeat_region	/rpt_family="L1PA6" 19461. 19550
repeat_region	/rpt_family="L2" 19562. 19591
repeat_region	/rpt_family="(TGAA)n" complement(20511. 20745)
repeat_region	/rpt_family="MIR" complement(20999. 21095)
repeat_region	/rpt_family="L2" 21284. 21520
repeat_region	/rpt_family="MLT1J1" 21949. 21991
repeat_region	/rpt_family="(TTTGG)n" 23647. 23713
repeat_region	/rpt_family="(CA)n" 24203. 24340
repeat_region	/rpt_family="MIR" complement(24837. 25022)
repeat_region	/rpt_family="MIR3" complement(25220. 25371)
repeat_region	/rpt_family="L2" 25505. 25811
repeat_region	/rpt_family="ERVL-F" 25815. 26206
repeat_region	/rpt_family="MLT2F" 26278. 26373
repeat_region	/rpt_family="MLT2F" 26395. 26663
repeat_region	/rpt_family="L1PAL6" 26664. 26953
repeat_region	/rpt_family="AluSx" 26954. 27438

Query Match 42.9%; Score 30; DB 9; Length 150883;
Best Local Similarity 72.2%; Pred. No. 19;
Matches 39; Conservative 0; Mismatches 15; Indels: 0

QY

3	TCCCTTCTCGGAAATGGAAACTGAATAACAGATCCCCTCTCGGAAATGGAAA	56

Db

42450	TCCCGTCACGAAATGGCTACTGAAATTAGGACCCTCTTGSCAGTTGTAA	42397

RESULT 11

AL591711	AL591711	197743 bp	DNA	linear	ROD 25-MAY-2002
LOCUS	Mouse DNA sequence from clone RP23-19L12 on chromosome 2, complete sequence.				
DEFINITION	AL591711				
ACCESSION	AL591711.22	GI:21217770			
VERSION	HTG.				
KEYWORDS	house mouse.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 197743)				
TITLE	Johnson,C.				
JOURNAL	Direct Submission				
	Submitted (25-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,				
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:				
	humuq@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk				
	On May 27, 2002 this sequence version replaced gi:21068504.				
COMMENT	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.				
	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate				

SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 196360)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissegh,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
 Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstock,G. and Gibbs,R.

Direct Submission
 Unpublished
 2 (bases 1 to 196360)
 AUTHORS Worley,K.C.

Direct Submission
 Submitted (24-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 196360)
 AUTHORS Worley,K.C.

Direct Submission
 Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 11, 2002 this sequence version replaced gi:17973146.

COMMENT -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GHFL
 Center clone name: CH230-64F11
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329

Consensus quality: 130985 bases at least Q40
 Consensus quality: 137495 bases at least Q30
 Consensus quality: 142650 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 78 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1
 1005: contig of 1005 bp in length
 1105: gap of unknown length
 1106: contig of 1015 bp in length
 2121: gap of unknown length
 2221: contig of 1051 bp in length
 3272: gap of unknown length
 3372: contig of 1571 bp in length
 4943: gap of unknown length
 5042: contig of 1653 bp in length
 6696: gap of unknown length
 6795: contig of 1687 bp in length
 8482: gap of unknown length
 8483: gap of unknown length
 8583: contig of 1029 bp in length
 9612: gap of unknown length
 9712: contig of 1091 bp in length
 10802: contig of 1091 bp in length
 10902: gap of unknown length
 11944: contig of 1042 bp in length
 12044: gap of unknown length
 13423: contig of 1379 bp in length
 13523: gap of unknown length
 14596: contig of 1073 bp in length
 14597: gap of unknown length
 16601: contig of 1905 bp in length
 16701: gap of unknown length
 17907: contig of 1206 bp in length
 18007: gap of unknown length
 19613: contig of 1606 bp in length
 19713: gap of unknown length
 20826: contig of 1113 bp in length
 20926: gap of unknown length
 22558: contig of 1632 bp in length
 22658: gap of unknown length
 24413: contig of 1755 bp in length
 24513: gap of unknown length
 25973: contig of 1460 bp in length
 26073: gap of unknown length
 27357: contig of 1284 bp in length
 27457: gap of unknown length
 28672: contig of 1215 bp in length
 28772: gap of unknown length
 30074: contig of 1302 bp in length
 30174: gap of unknown length
 31717: contig of 1543 bp in length
 31817: gap of unknown length
 33178: contig of 1361 bp in length
 33278: gap of unknown length
 34465: contig of 1187 bp in length
 34565: gap of unknown length
 37040: contig of 2475 bp in length
 37140: gap of unknown length
 38197: contig of 1057 bp in length
 38297: gap of unknown length
 40231: contig of 1934 bp in length
 40331: gap of unknown length
 42055: contig of 1724 bp in length
 42155: gap of unknown length
 43274: contig of 1119 bp in length
 43374: gap of unknown length
 45875: contig of 2501 bp in length

Insert size: 158298; sum-of-contigs

Insert size: 158
Quality coverage:

NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

11115: contig of 1115 bp in length	1116 1215: gap of 100 bp	1216 2414: contig of 1199 bp in length	2415 2514: gap of 100 bp	2515 3561: contig of 1047 bp in length	3562 3661: gap of 100 bp	3662 4929: contig of 1288 bp in length	4930 5029: gap of 100 bp	5030 6039: contig of 1010 bp in length	6040 6139: gap of 100 bp	6140 7469: contig of 1330 bp in length	7470 7569: gap of 100 bp	7570 9736: contig of 2167 bp in length	9737 9836: gap of 100 bp	9838 12165: contig of 2329 bp in length	12166 12265: gap of 100 bp	12266 12623: contig of 358 bp in length	12624 12723: gap of 100 bp	12724 14465: contig of 1742 bp in length	14466 14565: gap of 100 bp	14566 15963: contig of 1398 bp in length	15964 16063: gap of 100 bp	16064 18074: contig of 2011 bp in length	18075 18174: gap of 100 bp	18175 21150: contig of 2976 bp in length	21151 21250: gap of 100 bp	21251 22852: contig of 1602 bp in length	22853 22952: gap of 100 bp	22953 25375: contig of 2423 bp in length	25376 25475: gap of 100 bp	25476 27959: contig of 2484 bp in length	27960 28059: gap of 100 bp	28060 30399: contig of 2340 bp in length	30400 30499: gap of 100 bp	30500 33415: contig of 3816 bp in length	33416 37936: contig of 3521 bp in length	37937 38036: gap of 100 bp	38037 40123: contig of 2086 bp in length	40123 40223: gap of 100 bp	40223 43895: contig of 3673 bp in length	43896 43995: gap of 100 bp	43996 49451: contig of 5456 bp in length	49452 49551: gap of 100 bp	49552 54107: contig of 4556 bp in length	54108 54207: gap of 100 bp	54208 60101: contig of 5894 bp in length	60102 60201: gap of 100 bp	60202 64961: contig of 4760 bp in length	64962 65061: gap of 100 bp	65062 75980: contig of 10919 bp in length	75981 76080: gap of 100 bp	76081 89332: contig of 13252 bp in length	89333 89432: gap of 100 bp	89433 104747: contig of 15315 bp in length	104748 104847: gap of 100 bp	104848 120555: contig of 15708 bp in length	120556 120655: gap of 100 bp	120656 161198: contig of 40543 bp in length.
------------------------------------	--------------------------	--	--------------------------	--	--------------------------	--	--------------------------	--	--------------------------	--	--------------------------	--	--------------------------	---	----------------------------	---	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	--	----------------------------	---	----------------------------	---	----------------------------	--	------------------------------	---	------------------------------	--

Location/Qualifiers

```
1. .161198
/organism="Homo sapiens"
/db xref="taxon:9606"
```

FEATURES
source

```

/misc_feature /clone="RP11-71A1"
/misc_feature /clone_lib="RPC1-i1 Human Male BAC"
/misc_feature 1..1115
/misc_feature 1216..2414
/misc_feature /note="assembly_fragment"
/misc_feature 2515..3561
/misc_feature /note="assembly_fragment"
/misc_feature 3662..4929
/misc_feature /note="assembly_fragment"
/misc_feature 5030..6039
/misc_feature /note="assembly_fragment"
/misc_feature 6140..7469
/misc_feature /note="assembly_fragment"
/misc_feature 7570..9736
/misc_feature /note="assembly_fragment"
/misc_feature 9837..12165
/misc_feature /note="assembly_fragment"
/misc_feature 12266..12623
/misc_feature /note="assembly_fragment"
/misc_feature clone_end.T7
/misc_feature vector_side:right
/misc_feature 12724..14465
/misc_feature /note="assembly_fragment"
/misc_feature 14566..15963
/misc_feature /note="assembly_fragment"
/misc_feature 16064..18074
/misc_feature /note="assembly_fragment"
/misc_feature 18175..21150
/misc_feature /note="assembly_fragment"
/misc_feature 21251..22852
/misc_feature /note="assembly_fragment"
/misc_feature 22953..25375
/misc_feature /note="assembly_fragment"
/misc_feature 25476..27959
/misc_feature /note="assembly_fragment"
/misc_feature 28080..30399
/misc_feature /note="assembly_fragment"
/misc_feature 30500..34315
/misc_feature /note="assembly_fragment"
/misc_feature 34416..37936
/misc_feature /note="assembly_fragment"
/misc_feature 38037..40122
/misc_feature /note="assembly_fragment"
/misc_feature clone_end.SP6
/misc_feature vector_side:right
/misc_feature 40223..43895
/misc_feature /note="assembly_fragment"
/misc_feature 43996..49451
/misc_feature /note="assembly_fragment"
/misc_feature 49552..54107
/misc_feature /note="assembly_fragment"
/misc_feature 54208..60101
/misc_feature /note="assembly_fragment"
/misc_feature 60202..64961
/misc_feature /note="assembly_fragment"
/misc_feature 65062..75980
/misc_feature /note="assembly_fragment"
/misc_feature 76081..89332
/misc_feature /note="assembly_fragment"
/misc_feature 89433..104747
/misc_feature /note="assembly_fragment"
/misc_feature 104848..120555
/misc_feature /note="assembly_fragment"
/misc_feature 120656..161198
/misc_feature /note="assembly_fragment"
COUNT 40700 a 38352 c 37941 g 41304 t

```

Query Match 45.1%; Score 31.6; DB 2; Length 161198;
Best Local Similarity 74.1%; Pred. NO. 5.9;
Matches 40; Conservative 0; Mismatches 14; Indels 0;

0.

* 27254 28110: contig of 857 bp in length
* 28111 28210: gap of 100 bp
* 28211 29057: contig of 847 bp in length
* 29058 29157: gap of 100 bp
* 29158 30015: contig of 858 bp in length
* 30016 30115: gap of 100 bp
* 30116 30959: contig of 844 bp in length
* 30960 31059: gap of 100 bp
* 31060 31892: contig of 833 bp in length
* 31893 31992: gap of 100 bp
* 31993 32825: contig of 833 bp in length
* 32826 32925: gap of 100 bp
* 32926 33786: contig of 861 bp in length
* 33787 33886: gap of 100 bp
* 33887 34730: contig of 844 bp in length
* 34731 34830: gap of 100 bp
* 34831 35686: contig of 856 bp in length
* 35687 35786: gap of 100 bp
* 35787 35886: contig of 856 bp in length
* 35887 36642: contig of 846 bp in length
* 36643 36742: gap of 100 bp
* 36743 37588: contig of 846 bp in length
* 37589 37688: gap of 100 bp
* 37689 38583: contig of 895 bp in length
* 38584 38683: gap of 100 bp
* 38684 39538: contig of 855 bp in length
* 39539 39638: gap of 100 bp
* 39639 40465: contig of 827 bp in length
* 40466 40565: gap of 100 bp
* 40566 41398: contig of 833 bp in length
* 41399 41498: gap of 100 bp
* 41499 42350: contig of 852 bp in length
* 42351 42450: gap of 100 bp
* 42451 43323: contig of 873 bp in length
* 43324 43423: gap of 100 bp
* 43424 44274: contig of 851 bp in length
* 44275 44374: gap of 100 bp
* 44375 45243: contig of 869 bp in length
* 45244 45343: gap of 100 bp
* 45344 46198: contig of 855 bp in length
* 46199 46298: gap of 100 bp
* 46299 47210: contig of 912 bp in length
* 47211 47310: gap of 100 bp
* 47311 48149: contig of 839 bp in length
* 48150 48249: gap of 100 bp
* 48250 49095: contig of 846 bp in length
* 49096 49195: gap of 100 bp
* 49196 50025: contig of 830 bp in length
* 50026 50125: gap of 100 bp
* 50126 50955: contig of 830 bp in length
* 50956 51055: gap of 100 bp
* 51056 51889: contig of 834 bp in length
* 51890 51989: gap of 100 bp
* 51990 52860: contig of 871 bp in length
* 52861 52960: gap of 100 bp
* 52961 53793: contig of 833 bp in length
* 53794 53893: gap of 100 bp
* 53894 54738: contig of 845 bp in length
* 54739 54838: gap of 100 bp
* 54839 55682: contig of 844 bp in length
* 55683 55782: gap of 100 bp
* 55783 55640: contig of 858 bp in length
* 55641 56740: gap of 100 bp
* 56741 57576: contig of 836 bp in length
* 57577 57676: gap of 100 bp
* 57677 58527: contig of 851 bp in length
* 58528 58627: gap of 100 bp
* 58628 59468: contig of 841 bp in length
* 59469 59568: gap of 100 bp
* 59569 60400: contig of 832 bp in length
* 60401 60500: gap of 100 bp
* 60501 61390: contig of 890 bp in length
* 61391 61490: gap of 100 bp
* 61491 62346: contig of 856 bp in length

* 62347 62446: gap of 100 bp
* 62447 63310: contig of 864 bp in length
* 63311 63410: gap of 100 bp
* 63411 64262: contig of 852 bp in length
* 64263 64362: gap of 100 bp
* 64363 65192: contig of 830 bp in length
* 65193 65292: gap of 100 bp

Query Match 45.1%; Score 31.6; DB 2; Length 69861;
Best Local Similarity 74.1%; Pred. No. 5.7;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GAATGGAACAGAAATCAGATCCCTTCGGGAATGGAACAGAAATCAG 67
||||| ||||||| ||| ||||||| ||| ||||||| ||||||| ||| |||
DB 1327 GAATGGAACAGAAATCAGATCCCTTCGGGAATGGAATTCGAATCTTAG 1380

RESULT 3
AC015867
LOCUS Homo sapiens clone RP11-71A1, WORKING DRAFT SEQUENCE, 30 unordered
DEFINITION pieces.
ACCESSION AC015867
VERSION AC015867.2 GI:7249301
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE 1 (bases 1 to 161198)
JOURNAL Homo sapiens, clone RP11-71A1
REFERENCE 2 (bases 1 to 161198)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boquslavkiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S.,
Grant,G., Hago,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R.,
Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H.,
O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S.,
Tirrell,A., Vassiliev,H., Vo.A., Wheeler,J., Wu,X., Wyman,D., Ye.W.J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 16, 2000 this sequence version replaced gi:6446801.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L641
Center clone name: 71_A1
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Sequencing vector: Plasmid; n/a; 0.0 f% of reads
0.273224043715847Chemistry: Dye-primer-amersham; 9% of reads
Chemistry: Dye-terminator Big Dye; 91% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147496 bases at least Q40
Consensus quality: 154226 bases at least Q30
Consensus quality: 156653 bases at least Q20
Insert size: 137000; agarose-tp

gene expression with the option of cell growth control
 Patent: WO 0065074-A 2 02-NOV-2000;
 Gesellschaft für Biotechnologische Forschung mbH (GBF) (; DE)
 FEATURES
 source 1. .70
 Location/Qualifiers
 /organism="Mesocricetus sp."
 /db_xref="taxon:10038"
 /note="baby hamster kidney cells"
 BASE COUNT 25 a 15 c 15 g 15 t
 ORIGIN

Query Match 100.0%; Score 70; DB 6; Length 70;
 Best Local Similarity 100.0%; Pred. No. 2e-12;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCCCTTCTCGGAAATGGAACATGAAATCAGATCCCTTCGCGAAATGGAACATCA 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 GATCCCTTCTCGGAAATGGAACATGAAATCAGATCCCTTCGCGAAATGGAACATCA 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 61 AAATCAGATC 70
 |||||||||
 Db 61 AAATCAGATC 70

AC032033 69861 bp DNA linear HTG 03-APR-2000
 AC032033
 LOCUS Homo sapiens chromosome 18 clone RP11-392F20 map 18, LOW-PASS
 DEFINITION SEQUENCE SAMPLING.

AC032033
 VERSION AC032033.1 GI:7387453
 KEYWORDS HTG; HTGS-PHASE0.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, E.,
 Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campioniano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Willson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 Submitted (03-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L9119
 Center clone name: 392_F_20

 * NOTE: This record contains 74 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 848: contig of 848 bp in length
 849 948: gap of 100 bp
 949 1814: contig of 866 bp in length
 1815 1914: gap of 100 bp
 1915 2781: contig of 867 bp in length
 2782 2881: gap of 100 bp
 2882 3726: contig of 845 bp in length
 3727 3826: gap of 100 bp
 3827 4682: contig of 856 bp in length
 4683 4782: gap of 100 bp
 4783 5612: contig of 830 bp in length
 5613 5712: gap of 100 bp
 5713 6560: contig of 848 bp in length
 6561 6660: gap of 100 bp
 6661 7508: contig of 848 bp in length
 7509 7608: gap of 100 bp
 7609 8438: contig of 830 bp in length
 8439 8538: gap of 100 bp
 8539 9375: contig of 837 bp in length
 9376 9475: gap of 100 bp
 9476 10341: contig of 866 bp in length
 10342 10441: gap of 100 bp
 10442 11277: contig of 836 bp in length
 11278 11377: gap of 100 bp
 11378 12218: contig of 841 bp in length
 12219 12318: gap of 100 bp
 12319 13169: contig of 851 bp in length
 13170 13269: gap of 100 bp
 13270 13976: contig of 707 bp in length
 13977 14076: gap of 100 bp
 14077 14931: contig of 855 bp in length
 14932 15031: gap of 100 bp
 15032 15872: contig of 841 bp in length
 15873 15972: gap of 100 bp
 15973 16802: contig of 830 bp in length
 16803 16902: gap of 100 bp
 16903 17723: contig of 821 bp in length
 17724 17823: gap of 100 bp
 17824 18674: contig of 851 bp in length
 18675 18774: gap of 100 bp
 18775 19632: contig of 858 bp in length
 19633 19732: gap of 100 bp
 19733 20580: contig of 848 bp in length
 20581 20680: gap of 100 bp
 20681 21532: contig of 852 bp in length
 21533 21632: gap of 100 bp
 21633 22477: contig of 845 bp in length
 22478 22577: gap of 100 bp
 22578 23424: contig of 847 bp in length
 23425 23524: gap of 100 bp
 23525 24327: contig of 803 bp in length
 24328 24427: gap of 100 bp
 24428 25243: contig of 816 bp in length
 25244 25343: gap of 100 bp
 25344 26186: contig of 843 bp in length
 26187 26286: gap of 100 bp
 26287 27153: contig of 867 bp in length
 27154 27253: gap of 100 bp

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:16 ; Search time 385.676 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: us-09-980-548-2
Perfect score: 70
Sequence: 1 gatccctctcgggaatgg.....tggaaactgaaatcagatc 70

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				GenEmbl.:*							
1:	gb_ba:*	C 22	28.2	40.3	70522	2	AC130491	Homo sapi	AC130491	Homo sapi	
2:	gb_ba:*	C 23	28.2	40.3	78603	2	HS1188A21	Human DNA	AL117373	Human DNA	
3:	gb_htg:*	C 24	28.2	40.3	176774	2	AP002519	Homo sapi	AP002519	Homo sapi	
4:	gb_in:*	C 25	28	40.0	31976	9	AL596107	Human DNA	AL596107	Human DNA	
5:	gb_om:*	C 26	28	40.0	120724	9	CNS05TCR	Human chr	AL355837	Human chr	
6:	gb_ov:*	C 27	28	40.0	144223	9	AC062015	Homo sapi	AC062015	Homo sapi	
7:	gb_ov:*	C 28	28	40.0	145458	2	AC069591	Homo sapi	AC069591	Homo sapi	
8:	gb_pat:*	C 29	28	40.0	197659	9	AC009245	Homo sapi	AC009245	Homo sapi	
9:	gb_pat:*	C 30	27.8	39.7	125590	9	AC002382	Human BAC	AC002382	Human BAC	
10:	gb_pl:*	C 31	27.8	39.7	141905	2	AC110101	Rattus no	AC110101	Rattus no	
11:	gb_ro:*	C 32	27.8	39.7	161831	2	AC073078	Homo sapi	AC073078	Homo sapi	
12:	gb_sts:*	C 33	27.8	39.7	163567	2	AC120955	Rattus no	AC120955	Rattus no	
13:	gb_sy:*	C 34	27.8	39.7	180318	2	AC109991	Rattus no	AC109991	Rattus no	
14:	gb_un:*	C 35	27.8	39.7	186862	2	AC110483	Rattus no	AC110483	Rattus no	
15:	gb_vi:*	C 36	27.8	39.7	214617	2	AC127366	Mus muscu	AC127366	Mus muscu	
16:	em_ba:*	C 37	27.8	39.7	214934	2	AC068907	Mus muscu	AC068907	Mus muscu	
17:	em_fun:*	C 38	27.8	39.7	235079	2	AC124475	Mus muscu	AC124475	Mus muscu	
18:	em_hum:*	C 39	27.6	39.4	2352	3	DAHSOP70	X78403 D.auraria m	X78403 D.auraria m		
19:	em_in:*	C 40	27.6	39.4	136641	2	AC114327	Canis fam	AC114327	Canis fam	
20:	em_mu:*	C 41	27.6	39.4	168703	4	AC091759	Sus scrofa	AC091759	Sus scrofa	
21:	em_om:*	C 42	27.6	39.4	172884	2	AC113567	Canis fam	AC113567	Canis fam	
22:	em_or:*	C 43	27.6	39.4	184234	2	AC091758	Sus scrofa	AC091758	Sus scrofa	
23:	em_ov:*	C 44	27.6	39.4	184509	2	AC113237	Canis fam	AC113237	Canis fam	
24:	em_pat:*	C 45	27.4	39.1	26483	2	AC100441	Mus muscu	AC100441	Mus muscu	

ALIGNMENTS

RESULT 1
AX040910
LOCUS AX040910 70 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 2 from Patent WO0065074.
ACCESSION AX040910
VERSION AX040910.1 GI:11340532
KEYWORDS
SOURCE Mesocricetus sp.
ORGANISM Mesocricetus sp.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
Mesocricetus.
1. (bases 1 to 70)
Mueller,P., Geserick,C., Schroeder,K. and Hauser,H.
Promoter-transactivator system for inducible high-level mammalian

Pred. No. is the number of results predicted by chance to have a

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y., Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

Location/Qualifiers
1. .535

```
/clone_lib="RIKEN I
/cell_type="B16 F10
/note="pooled tissu
dev stage=16 days n
```

Sample	134 c	120 g
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		
61		
62		
63		
64		
65		
66		
67		
68		
69		
70		
71		
72		
73		
74		
75		
76		
77		
78		
79		
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90		
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		

Qy	1	GO
Db	111	GO

AAAGTCAGGAACAGAGAAAC
| | | | | | | |
CAAGGAAGTACAGAGAGGC

```
Qy      113 A    |  
        231 A    |  
  
Db      172 TT   |  
        291 TF   |  
  
Search completed  
Job time : 20
```

/note="Vector: pSPORT1; Site.1: SalI; Site.2: NotI: This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 134 a 127 c 98 g 97 t
ORIGIN
Query Match 21.9%; Score 59; DB 10; Length 456;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;
QY 1 GCTAGCTTAAGTACGCCATTTTCAGGCGTGGGAAATACATACTGAGATAGAGA 60
Db 22 GCTAGCTGCAATGCGCATTTTCAGGCGTGGGAAATACATACTGAGATAGAGA 81
QY 61 AGTTCAGATCAAGTTCAGGACAGAGAACAGAGAT-----ATGGCCAAA 108
Db 82 AGTTACAAGGAAGTTAGTTAAATAAATAAGGCTGAATACTGAGACAGAGCCAAA 141
QY 109 CAGGATATCTGTGT-AAGCAGTTCTCTGCCCGCTCAGGCGCCAGAGAGATTTGGAAC 164
Db 142 CAGGATATCTGTGTCAAGCACCTGGCCCGCTCAGGCGCCAGAGAGATTTGTTAC 198
RESULT 13
BB756171 506 bp mRNA linear EST 17-OCT-2001
LOCUS BB756171 RIKEN full-length enriched, melanocyte Mus musculus cDNA
DEFINITION clone G270072B14 3', mRNA sequence.
ACCESSION BB756171.1 GI:16202641
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 506)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okado,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
TITLE Unpublished (2001)
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic
COMMENT Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.
FEATURES
Location/Qualifiers
1..506
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G270072B14"
/clone_lib="RIKEN full-length enriched, melanocyte"
/cell_type="melanocyte"
/note="pooled tissues; (tissue_type=cerebellum, dev_stage=16 days neonate, sex=mixed), (tissue_type=cerebellum, dev_stage=0 day neonate, sex=mixed), (tissue_type=hippocampus, dev_stage=adult, sex=male), (tissue_type=whole body, dev_stage=9 days embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days embryo, sex=mixed)"
BASE COUNT 153 a 130 c 112 g 111 t
ORIGIN
Query Match 21.9%; Score 59; DB 10; Length 506;
Best Local Similarity 67.2%; Pred. No. 2e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;
QY 1 GCTAGCTTAAGTACGCCATTTTCAGGCGTGGGAAATACATACTGAGATAGAGA 60
Db 72 GCTAGCTCAGTACGCCATTTTCAGGCGTGGGAAATACATACTGAGATAGAGA 131
QY 61 AGTTCAGATCAAGTTCAGGACAGAGAACAGAGAT-----ATGGCCAAA 108
Db 132 AGTTACAAGGAAGTTAGTTAAATAAATAAGGCTGAATACTGAGACAGAGCCAAA 191
QY 109 CAGGATATCTGTGT-AAGCAGTTCTCTGCCCGCTCAGGCGCCAGAGAGATTTGGAAC 164
Db 192 CAGGATATCTGTGTCAAGCACCTGGCCCGCTCAGGCGCCAGAGAGATTTGTTAC 248
RESULT 14
BB753609 787 bp mRNA linear EST 25-SEP-2001
LOCUS BB753609 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5198973 5',
DEFINITION mRNA sequence.
ACCESSION BB753609
VERSION BB753609.1 GI:15745187
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 787)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takanashi, F., Takaku-Akahara, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muraatsu, M., and Hayashizaki, Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
S., Kawaj, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
Y., and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.

FEATURES source

Location/Qualifiers
1. .456
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="E860002A18"
/clone_lib="RIKEN full-length enriched, 8 cells embryo"
/cell_type="8 cells"
/dev_stage="8 cells embryo"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT) -Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

134 a 127 c 98 g 97 t

```

Query Match      21.9%  Score 59;  DB 10;  Length 456;
Best Local Similarity 67.2%;  Pred. No. 1.9e-05;
Matches 119;  Conservative 0;  Mismatches 45;  Indels 13;  Gaps 2;

OY 1  GCTAGCTTAAGTACGGCCATTTCGCAAGGCATGGGAAAAATACATACTGAGAATAGAGA 60
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 22  GCTAGCTGCGATACGGCCATTTCGCAAGGCATGAAAAGTACCAGAGCTGAGTCTCTCAA 81
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Qy	61	AGTTCAGATCAAGCTCAGGACAGAGAAACAGGAGAAT-----ATGGGCCAAA	108
Dd	82	AGTTACAAGGAAGTTTTAGTTAAAAAAAATAAAGCGCTGAATATACTGACACAGAGCCCAA	141
Qy	109	CAGGATATCTGTGCT-AAGCAGATTCTTGCCCGCTCAGGGCCAAGAACAGATTGGAAC	164
Dd	142	CAGGATATCTGTGCTCAAGCACCTTGCGCCCGCTCAGGGCCAAGACAGATGGTAC	198
RESULT	12		
LOCUS	BB733115		
DEFINITION	BB733115 RIKEN full-length enriched, 12 days embryo whole body Mus musculus cDNA clone E970035B21 3', mRNA sequence.	EST	linear EST 12-OCT-2001
ACCESSION	BB733115		
VERSION	BB733115.1	GI:16116390	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 456)		
AUTHORS	AKimura,T., ARAkawa,T., CARninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., HIRamoto,K., HiraoKa,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J.J., KoJima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shlraki,I., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Wachihi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki.Y.		
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.)		

JOURNAL
COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 1. .456
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="E970035821"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 whole body"
 /tissue_type="whole body"
 /dev_stage="12 days embryo"

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="E970035B21"
/clone_lib="RIKEN full-length enriched, 12 days embryo
whole body"
/tissue_type="whole body"
/dev_stage="12 days embryo"

```

transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATCTCGATTAAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT
ORIGIN

131 a 122 c 96 g 96 t

Query Match 21.9%; Score 59; DB 10; Length 445;
Best Local Similarity 67.2%; Pred. No. 1.9e-05;
Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;

QY 1 GCTAGCTTAAGTAAACGCCATTTTCGAAGGCATGGGAAAAATACATACTAGAGANTAGAGA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 12 GCTAGCTGCAATCAAGCCATTTTCGAAGGCATGAAAAGATTACCAGAGCTGAGTTCTCAA 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 61 AGTTTCAGATCAAGGTCAAGAACAGAGAACAGGAGAAAT-----ATGGGCCAAA 108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 72 AGTTACAAGGAAGTTTAGTTAAAAAATAAAGCTGAATAACTTGAGACAGAGGCCAAA 131
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 109 CAGGATATCTGTGGT-AAGCAGTTCTCGCCCGCTCAGGCGCCAAGAACAGTTGGAAC 164
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 CAGGATATCTGTGGTCAAGCACCTTGGCCCCGGCTCAGGCGCCAAGAACAGATGGTAC 188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BB670350

LOCUS BB670350 linear EST 05-OCT-2001

DEFINITION musculus cDNA clone 3322402123 3', mRNA sequence.

ACCESSION BB670350

VERSION BB670350.1 GI:15969571

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 455)

REFERENCE Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T., Ishii,Y., Ito,M., Kawai,J., Koizumi,K., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishikawa,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y., RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y., Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuiura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y., RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,

Y. and Hayashizaki, Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES	source	Location/Qualifiers
	1. .455	/organism="Mus musculus"
		/strain="C57BL/6J"
		/db_xref="taxon:10090"
		/clone="3322402123"
		/clone_lib="RIKEN full-length enriched, 17 days embryo head"
		/sex="mixed"
		/tissue_type="head"
		/dev_stage="17 days embryo"
		/lab_host="DH10B"
		/notes="Site.1: Sali; Site.2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATTCTCGAGTTAAATTAATCCGCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I".
BASE COUNT	134 a	126 c 98 g 97 t
ORIGIN		
Query Match	21.9%	Score 59; DB 10; Length 455;
Best Local Similarity	67.2%	Prod. No. 1.9e-05;
Matches 119; Conservative	0; Mismatches 45; Indels 13; Gaps 2;	
QY	1	GCTAGCTTAAGTACGCCATTTCGCAAGGATCGGAAATACTACTAGATAGAGA 60
Db	21	GCTAGCTCGATACGCCATTTTCGCAAGGATGAAAAGTACCAGAGCTGAGTCTCAA 80
QY	61	AGTTCAGATCAAGTCAAGGAACAGAGAGAGAGAT-----ATGGGCCCAA 108
Db	81	AGTTACAAGGAAGTTAGTTAAAAAATAAGGCTGAATACTAGACAGAGGCCAAA 140
QY	109	CAGATATCTGTGT-AAGCAGTTCCTGCCCGCTCAGGCGCCAAAGACAGTTGGAAC 164
Db	141	CAGGATATCTGTGTCAAGCACCCTGGCGCCCGCTCAGGCGCCAAAGACAGATGTGATC 197
RESULT 11		
BB726927		
LOCUS	BB726927	456 bp mRNA linear EST 12-OCT-2001
DEFINITION	BB726927	RIKEN full-length enriched, 8 cells embryo Mus musculus
ACCESSION	BB726927	cDNA clone E860021A18 3', mRNA sequence.
VERSION	BB726927.1	GI:161110202
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	1 (bases 1 to 456)	
	Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,	

Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Muramatsu, M., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

UNIGEN
 : COUNT 130 a 120 c 93 g 96 t
 Query Match 21.9%; Score 59; DB 10; Length 439;
 Best Local Similarity 67.2%; Pred. No. 1.9e-05;
 Matches 119; Conservative 0; Mismatches 45; Indels 13; Gaps 2;
 QY 1 GCTAGCTTAAGTAAGCCATTTTGCAGGCATGGGAAATAATACATACCTGAGAAATAGAGA 60
 Db 6 GCTAGCTGAGTAAGCCATTTTGCAGGCATGGGAAATAATACATACCTGAGAAATAGAGA 65
 QY 61 AGTTTCAGATCAGGTCAGCAACAGAGAAACAGAGAAAT-----ATGGGCCCAA 108
 Db 66 AGTTTCAAGGAAGTTAGTTAAATAAATAAAGGCTGAATAATCTAGACAGAGGCCAA 125
 QY 109 CAGGATATCTGTGGT-AAGCAGCTTCTGCCCGCTCAGGCCCAAGAACACAGTTGGTAC 164
 Db 126 CAGGATATCTGTGGTCAAGACCTGGGCCCGCTCAGGCCCAAGAACACAGATGGTAC 182

RESULT 9

BB670654
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

BB670654 445 bp mRNA linear EST 05-OCT-2001
 BB670654 RIKEN full-length enriched, 17 days embryo head Mus
 musculus cDNA clone 3322404N14 3', mRNA sequence.
 BB670654
 BB670654.1 GI:15969875
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hiramoto, K., Hiraoka, T., Konno, H., Kouda, M., Matsuyama, T.,
 Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
 Nakamura, M., Nishikawa, K., Numata, K., Numata, R., Okazaki, Y., Okada, T.,
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
 Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa
 A., Takahashi, F., Takaku-Akashira, S., Tanaka, T., Tomaru, A., Toya, T.,
 Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y., et al.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001)
 Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.
 e mouse tissues.

FEATURES
 Location/Qualifiers
 Source

1. 445
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="3322404N14"
 /clone_lib="RIKEN full-length enriched, 17 days embryo
 head"
 /sex="mixed"
 /tissue_type="head"
 /dev_stage="17 days embryo"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RIKEN full-length enriched, mammary gland
RCB-0527 Jyg-MC(B) CDNA"
/tissue_type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
/notes="pooled cell lines ; (cell_line=CRL-1751 WEHI 164),
(cell_line=CRL-2116 JC), (cell_line=RCB-0035 WEHI-3),
(cell_line=RCB-0464 Meth-A), (cell_line=RCB-0345 OHTA),
(cell_line=RCB-0559 K-1 F1), (cell_line=RCB-1283 B16
melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
), (cell_type=Itydig cells, cell_line=CRL-2065 MLTC-1),
(cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
(tissue_type=bladder, cell_line=RCB-0544 MEF-2),
(tissue_type=bone marrow, cell_type=stroma cell,
cell_line=CRL-2028 SR-4987), (tissue_type=colon,
cell_line=RCB-0549 Cle-H3), (tissue_type=kidney,
cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13, 20-3B3
), (strain=C3H, tissue_type=brain, cell_line=CRL-1443
B3H1).

```

```

BASE COUNT      133 a      130 c      98 g      85 t
ORIGIN
Query Match      22.8%; Score 61.6; DB 10; Length 446;
Best Local Similarity 63.2%; Pred. No. 3.9e-06;
Matches 115; Conservative 0; Mismatches 59; Indels 8; Gaps 1;

QY 1 GCTAGCTTAAGTACGGCCATTTTGCAGGCGATGGGAAAAATACATACTGAGAATAGAGA 60
||||| ||| ||||||||| ||||||||| ||||||||| ||| ||||| |
Db 25 GCTAACTCCAGTAACGGCCATCTTGCAGGCGATGGGAAAAATACCAGAGCTGATGTTCTCAG 84
||||| ||| ||||||||| ||||||||| ||||||||| ||| ||||| |
QY 61 AGTTCATCAAGCTCAGGACAGAGAACAGGAGATAT-----GGCCCAACAGG 112
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 85 AAAACAAGAACACAGGAAGTACAGAGGCTTGGAAAGTACCGGGAGCTAGGGCCCAACAGG 144
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 ATATCTGTGTAGTCAGTTCCTGCCCCGCTCAGGGCCCAAGAACAGTGTGGACACAGGAAAT 172
||||| ||| ||||||||| ||||||||| ||||||||| ||| ||||| |
Db 145 ATATCTGTGTCAACCACTAGGGGCCCGCCAGGCCCAAGAACAGATGGTCCCCAGAAAT 204
||||| ||| ||||||||| ||||||||| ||||||||| ||| ||||| |
QY 173 TG 174
|
Db 205 AG 206

```

RESULT 7	AW750456	284 bp	linear	EST 28-APR-2000
LOCUS	RC5-RT0604-260100-011-c11	BT0604	Homo sapiens	cdna, mRNA sequence.
DEFINITION	AW750456			
ACCESSION	AW750456.1	GI:7665388		
VERSION				
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 284)			
	Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)
MEDLINE	20202663			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,			

```

Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LIIC Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC56t2-RC5-BT0604-
260100-011-cl18t3-2000-01-286t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 283.
Location/Qualifiers
1. .284
;
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0604"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
FEATURES
source

```

BASE COUNT	45 a	76 c	65 g	98 t		
ORIGIN						
Query Match	22.0%;	Score 59.4;	DB 10;	Length 284;		
Best Local Similarity	67.4%;	Pred. No. 1.4e-05;				
Matches 118;	Conservative 0;	Mismatches 46;	Indels 11;	Gaps 2;		
QY	1	GCTAGCTTAAGTACGCCATTTTGC	CAAGGCATGGGAAAAATACATAACTGAGATAGAGA	60		
Db	220	GCTAGCTGCATACGCCATTTTGC	CAAGGCATGAAAAAGTACCAGAGCTGAGTTCTCAAA	161		
QY	61	AGTTCAGATCAAGTCAAGGACAGAGAAC	-----AGGAGATATATGGGCCCAACA	110		
Db	160	AGTCACAAGGAAGTTAGTTAAAGAAAT	AAGGCTGAACAAAACCTGGGACAGGGGCCAAACA	101		
QY	111	GGATATCTGTGGT-AAGCAGTTCTCTG	CCCCCGCTCAGGCCCAAGAACACAGTTGGAC	164		
Db	100	GGATATCTGTGGTCGACACCTGGG	CCCCCGCTCAGGGCCAAAGAACAGATGGGTAC	46		
RESULT 8						
BB733109						
LOCUS	BB733109	RIKEN full-length enriched, 12 days embryo whole body Mus	439 bp	mRNA	linear	EST 12-OCT-2001
DEFINITION	musculus cDNA clone E370034P16 3', mRNA sequence.					

SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 439)
AUTHORS	Hayakawa, T., Hiramoto, K., Carninci, P., Furuno, M., Hanagaki, T., Y. Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, K., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)
JOURNAL	Unpublished (2001).
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222

LOCUS BF101077 1240 bp mRNA linear EST 19-OCT-2000
 DEFINITION 601754641F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3983531 5',
 mRNA sequence.
 ACCESSION BF101077
 VERSION BF101077.1 GI:10883603
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 1240)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-re@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LAM9184 row: b column: 12
 High quality sequence stop: 580.
 Location/Qualifiers
 1..1240
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="IMAGE:3983531"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="10 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 472 a 310 c 308 g 149 t
 ORIGIN
 Query Match 23.2%; Score 62.6; DB 12; Length 1240;
 Best Local Similarity 58.6%; Pred. No. 2.5e-06;
 Matches 120; Conservative 0; Mismatches 44; Indels 11; Gaps 2;
 1 GCTAGCTTAAGTACCCATTTGCAAGGCATCGGAAATACATACTAGATAGAGA 60
 Db 148 GCTAGCTGAGTAACGCCATTTTGCAGGCATGAAAGATACCAGAGCTGAGTTCTCAA 207
 QY 61 AGTTTCAGATCAAGGTCAGGAACAGAGAAAC-----AGGAGAAATATGGGCCAAACA 110
 Db 208 AGTTACAAGGAAGTTAGTTAAGATTAGGCTGAATTAATACGGGACAGGGGCCAAACA 267
 QY 111 GGATATCTGTGT-AAGCAGTTCTCTCGCCCGCTCAGGGCCCAAGAACAGATGGGAAC 164
 Db 268 GGATATCTGTGTCAAGACACCTGGGCGCCGCTCAGGGCCCAAGAACAGATGGGTAC 322
 RESULT 4
 AZ804164/c
 LOCUS AZ804164 666 bp DNA linear GSS 16-FEB-2001
 DEFINITION 2M0064J24R Mouse 10kb plasmid UUGCIM library Mus musculus genomic
 clone UUGC2M0064J24 R, DNA sequence.
 ACCESSION AZ804164
 VERSION AZ804164.1 GI:12956487
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 REFERENCE 1 (bases 1 to 666)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A., and Wright, D., Weiss, R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0064 row: J column: 24
 Seq primer: CACACAGGAACACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 666.
 Location/Qualifiers
 1..666
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_lib="UUGC2M0064J24"
 /clone_lib="Mouse 10kb plasmid UUGCIM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gil4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
 BASE COUNT 141 a 166 c 148 g 211 t
 ORIGIN
 Query Match 22.9%; Score 61.8; DB 17; Length 666;
 Best Local Similarity 58.1%; Pred. No. 3.7e-06;
 Matches 155; Conservative 0; Mismatches 97; Indels 15; Gaps 2;
 1 GCTAGCTTAAGTACGCCATTTTGCAGGCATGAAAGATACCAGAGCTGAGTTCTCAA 276
 Db 335 GCTAGCTGCGATACGCCATTTTGCAGGCATGAAAGATACCAGAGCTGAGTTCTCAA 276
 QY 61 AGTTTCAGATCAAGGTCAGGAACAGAGAAATATGGGCCAAACAGATATCTGT 120
 Db 275 AGTTTACAGGAAGTTAGTTAAGATTAAAGATTAAAGTTAAAGTTAAAGTTAAAGTT 219
 QY 121 GGTAAAGCAGTTCTCGCCCGCTCAGGGCCCAAGAACAGTTGGAACAGGAGAAATGGGCCAA 180
 Db 218 AGGACAAGGCCCAAGAACAGGATGGTGCCAC-----CTGGGCCCAAGGATCAAG 171
 QY 181 ACAGGATATCTGTGGTAAGCAGTTCTCGCCCGCTCAGGGCCCAAGAACAGATGGTCCCA 240
 Db 170 GCCAAGAACCGGTGTACCCACCTGGGCCCGCTCAGGGCCCAAGAACAGATGGTACCA 111
 QY 241 GATCGCGTCCGCCCTCAGCAGTTTCT 267
 Db 110 GATAAGCGGGAACCAAGCAACAGTTTCT 84

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh

M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,

Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, T., Tanaka, T., Matsura

, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and

Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara

, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp) for

further details.

e mouse tissues.

BASE COUNT 172 a 137 c 117 g 112 t

ORIGIN

Query Match 23.4%; Score 63.2; DB 10; Length 538;

Best Local Similarity 63.7%; Pred. No. 1.5e-06;

Matches 116; Conservative 0; Mismatches 58; Indels 8; Gaps 1;

QY 1 GCTAGCTTAAGTAACCCATTTGCAAGCATGGGAAATACTGAGAAATAGAGA 60

Db 116 GCTAACTGCAGTAACCCATCTTGAAGCATGGGAAATACTGAGAAATAGAGA 175

QY 61 AGTTCAGATCAAGGTCAGACAGAAACAGGAGAAATAT-----GGGCAACACAGG 112

176 AAAACACAGACAGGAAGTACAGAGAGCTGGAAGTACCGGGAGTAGGGCCAAACAGG 235

QY 113 ATATCTGTGGTAAGCAGTTCTCCCGCTCAGGCGCAAGACAGTTGCAACAGGAGAAAT 172

Db 236 ATATCTGTGGTCAAGCACTAGGCGCGCGCCAGGCGCAAGACAGATGTCCTCCAGAAAT 295

QY 173 TG 174

Db 296 AG 297

RESULT 2

AZ984553 616 bp DNA linear GSS 27-APR-2001

LOCUS 2M0266C09F Mouse 10kb plasmid UUGC2M library Mus musculus genomic

DEFINITION clone UUGC2M0266C09 F, DNA sequence.

ACCESSION AZ984553

VERSION AZ984553.1 GI:13855780

KEYWORDS house mouse.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 616)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Rellly

, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.

and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dduenne@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0266 row: C column: 09

Seq primer: CTTGTAAACGACGCCAGT

Class: plasmid ends

High quality sequence stop: 616.

Location/Qualifiers

1. 616

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUGC2M0266C09"

/clone_lib="Mouse 10kb plasmid UUGC2M library"

/sex="Female"

/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"

/note="Vector: PWD42nv: Purified genomic DNA from M.

musculus C57BL/6J (female) was obtained from the Jackson

Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA

was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4

polynucleotide kinase. Adaptor oligonucleotides were

ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel

electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gil4732114|gb|AF129072.1), a copy-number

inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and

purified. The sheared, adapted mouse DNA was annealed to

adapted vector DNA, and transformed into

chemically-competent E. coli XL10-Gold (Stratagene) cells

and selected for ampicillin resistance."

BASE COUNT 180 a 151 c 141 g 144 t

ORIGIN

Query Match 23.2%; Score 62.6; DB 17; Length 616;

Best Local Similarity 68.6%; Pred. No. 2.2e-06;

Matches 120; Conservative 0; Mismatches 44; Indels 11; Gaps 2;

QY 1 GCTAGCTTAAGTAACCCATTTGCAAGCATGGGAAATACTGAGAAATAGAGA 60

Db 225 GCTAGCTGCAGTAACGCCATTTTGAAGCATGAAAAAGTACCAGAGCTGAGTCTCTCAA 284

QY 61 AGTTCAGATCAAGGTCAGACAGAAAC-----AGGAGAAATATGGGCCAAACA 110

Db 285 AGTTACAGGAAGTTAGTTAAAGATAAGGCTGAATAATACCGGACAGGCCAAACA 344

QY 111 GGATATCTCTGGT-AAGCAGTTCTCTCCCGCTCAGGCGCAAGACAGTTGGAGAC 164

Db 345 GGATATCTCTGGTCAAGCACCTGGGCGCCCGCTCAGGCGCAAGACAGATGGTAC 399

RESULT 3

BF101077

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:43:11 ; Search time 2036.09 Seconds
(without alignments)
2147.640 Million cell updates/sec

Title: US-09-980-548-1
Perfect score: 270
Sequence: 1 gctagtttaagtaagccat.....cgccctcagcagttttctaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

-eached: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	EST:*
1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	gb_estl:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63.2	23.4	538	10	BB751692
2	62.6	23.2	616	17	AZ984553
3	62.6	23.2	1240	12	BF101077
4	61.8	22.9	666	17	AZ804164
5	61.6	22.8	444	10	BB750356
6	61.6	22.8	446	10	BB835008

C	7	59.4	22.0	284	10	AW750456
	8	59	21.9	439	10	BB733109
	9	59	21.9	445	10	BB670654
	10	59	21.9	455	10	BB670350
	11	59	21.9	456	10	BB726927
	12	59	21.9	456	10	BB733115
	13	59	21.9	506	10	BB756171
	14	58.8	21.8	787	13	BI753609
	15	58.6	21.7	535	10	BB762345
	16	58.6	21.7	537	10	BB762595
C	17	58.6	21.7	584	13	BI794183
	18	58.6	21.7	596	13	BI677171
	19	58.6	21.7	644	10	BB795130
	20	58.6	21.7	649	10	BB795436
	21	58.6	21.7	650	10	BB794939
	22	58.6	21.7	660	10	BB794906
	23	58.6	21.7	662	10	BB795586
	24	58.6	21.7	663	10	BB794704
	25	58.6	21.7	666	10	BB795529
	26	58.6	21.7	680	10	BB794600
	27	58.6	21.7	680	10	BB794643
	28	58.6	21.7	687	10	BB794760
	29	58.6	21.7	688	10	BB794665
	30	58.6	21.7	691	10	BB794587
	31	57.6	21.3	209	9	AA666591
	32	57.6	21.3	209	9	AA146207
	33	57.6	21.3	210	9	AA655589
	34	57.6	21.3	210	9	AA655895
	35	57.6	21.3	210	9	AA684134
	36	57.6	21.3	210	9	AA122917
	37	57.6	21.3	269	9	AA139125
	38	57.6	21.3	444	10	BB831589
	39	57.6	21.3	426	12	BF162178
	40	57	21.1	726	10	BB786738
	41	57	21.1	435	10	BB821235
	42	57	21.1	444	10	BB788783
	43	57	21.1	449	10	BB815029
	44	57	21.1	461	10	BB833539
	45	57	21.1	525	10	BB762415

ALIGNMENTS

RESULT 1	BB751692	538 bp	mRNA	linear	EST 16-OCT-2001
LOCUS	BB751692	RIKEN full-length enriched, melanocyte	Mus musculus	CDNA	
DEFINITION	Clone G270003K01 3', mRNA sequence.				
ACCESSION	BB751692				
VERSION	BB751692.1	GI:16155928			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 538)				
AUTHORS	Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.				
TITLE	RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				

CC proteins in mammalian tissue culture host cells, including rat fibroblast
CC cells, bovine kidney cells and human kidney cells. The present sequence
CC is cytomegalovirus (CMV) MN14 vector used in the invention. The vector
CC comprises the following elements: CMV promoter, MN14 antibody heavy
CC chain signal peptide, MN14 antibody heavy chain, encephalomyocarditis
CC virus (ECMV) IRES -bovine alpha-lactalbumin signal peptide, MN14 antibody
CC light chain and 3' moloney murine leukemia virus (MoMuLV) LTR.

XX Sequence 4207 BP; 1030 A; 1202 C; 1058 G; 917 T; 0 other;

Query Match 75.1%; Score 202.8; DB 24; Length 4207;
Best Local Similarity 95.6%; Pred. No. 1.1e-55;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

Oy 1 GCTAGCTTAAGTAAGCCCATTTTGGCAAGGCATGGGAAATAATACATACTAGCAATAGAGA 60

Db 3644 GCTAGCTTAAGTAAGCCCATTTTGGCAAGGCAT-GGAAAAATACATACTAGCAATAGAGA 3702

Oy 61 AGTTCAGATCAAGTTCAGGAACAGA-GAAACAGGAGAAATATGGCCCAACAGGATATCTG 119

Db 3703 AGTTCAGATCAAGTTCAGGAACAGATGGAAACAGCTGAATATGGCCCAACAGGATATCTG 3762

Oy 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCCAAGAACAGTTGGAAACAGGAGAAAT-TGGGC 177

Db 3763 TGGTAAGCAGTTCCTGCCCCGCTCAGGGCCCAAGAACAGATGGAACAGCTGAATATGGGC 3822

Oy 178 CAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCCAAGAACAGATGGTC 236

Db 3823 CAACAGGATATCTGTGGTAAGCAGTTCCTGCCCCGCTCAGGGCCCAAGAACAGATGGTC 3882

Oy 237 CCCAGATCGGTCGCCGCTCAGCAGTTTCTAGA 270

Db 3883 CCCAGATCGGTCGCCGCTCAGCAGTTTCTAGA 3916

Search completed: March 5, 2003, 22:49:51
Job time : 277.478 secs

two integrated integrating vectors. The integrating vectors comprise at least one exogenous gene operably linked to a promoter. The host cell is useful for producing a desired protein and for comparing protein functions. The host cell comprises a reporter gene which is from gene fluorescent protein, luciferase, beta-galactosidase and beta-lactamase, and the assaying step further comprises detecting a signal from the reporter gene. The desired protein includes proteins for pharmaceutical, industrial, diagnostic and other purposes. The host cells is useful for indirectly detecting the expression of a desired protein, comprising providing the host cell transfected with a vector encoding a polycistronic sequence comprising a signal protein and a desired protein operably linked by an internal ribosome entry site (IRES), and culturing the host cell under suitable conditions so that the signal protein and the desired protein is produced, where the presence of signal protein indicates the presence of desired protein. The present sequence is cytomagalovirus (CMV) MN14 vector used in the invention. The vector comprises the following elements: CMV promoter, MN14 antibody heavy chain signal peptide, MN14 antibody heavy chain, encephalomyocarditis virus (ECMV) IRES -bovine alpha-lactalbumin signal peptide, MN14 antibody light chain and 3' moloney murine leukemia virus (MoMuLV) LTR.

Sequence 4207 BP; 1030 A; 1202 C; 1058 G; 917 T; 0 other;

very Match 75.1%; Score 202.8; DB 24; Length 4207;
Jest Local Similarity 95.6%; Pred. No. 1.1e-55;
Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

Qy 1 GCTAGCTTAAGTAACGCCATTTTGGCAAGCATGGGAAATAATACATACTGAGAATAGAGA 60
Dy 3644 GCTAGCTTAAGTAACGCCATTTTGGCAAGCAT-GGAAAAAATACATACTGAGAATAGAGA 3702

Qy 61 AGTTCAGATCAAGTCCAGAACAGAG-AGAACAGAGAGATATGGCCAAACAGGATATCTG 119
Dy 3703 AGTTCAGATCAAGTCCAGAACAGAGATATGGCCAAACAGGATATCTG 3762

Qy 120 TGGTAAGCAGTTCTCTGCCCC-GCTCAGGCGCCAGAACAGTGGAGCAGGAGAT-TGGGC 177
Dy 3763 TGGTAAGCAGTTCTCTGCCCCGCTCAGGCGCCAGAACAGATGGAACAGTGGGC 3822

Qy 178 CAACAGCATATCTGTGTAAGCAGTTCTGCCCC-GCTCAGGCGCCAGAACAGATGTC 236
Dy 3823 CAACAGCATATCTGTGTAAGCAGTTCTGCCCCGCTCAGGCGCCAGAACAGATGTC 3882

Qy 237 CCCAGATCGGTCGCCGCCCTCAGCAGTTTCTAGA 270
Dy 3883 CCCAGATCGGTCGCCGCCCTCAGCAGTTTCTAGA 3916

RESULT 15
AAD28306
ID AAD28306 standard; DNA; 4207 BP.

AAD28306;
22-APR-2002 (first entry)
Cytomegalovirus (CMV) MN14 vector.

Regulatory element; vector; erythropoietin; growth hormone; insulin; immunoglobulin; bone morphogenetic protein; interferon; interleukin; superoxide dismutase; T-cell receptor; surface membrane protein; CMV; viral antigen; transport protein; addressin; regulatory protein; ECMV; encephalomyocarditis virus; internal ribosome entry site; IRES; bovine; cytomagalovirus; moloney murine leukemia virus; MoMuLV; chimeric; alpha-lactalbumin; promoter; ds.

OS Chimeric - Encephalomyocarditis virus.
OS Chimeric - Bos sp.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Moloney murine leukemia virus.
OS Chimeric - Unidentified.

Key Location/Qualifiers

promoter 1..812
FT /tag= a
FT /note= "CMV promoter/enhancer"
FT misc_signal 853..855
FT /tag= b
FT /note= "MN14 antibody heavy chain gene signal peptide
FT start codon"
FT 2257..2259
FT /tag= c
FT /note= "MN14 antibody heavy chain gene start codon"
FT misc_feature 2271..2846
FT /tag= d
FT /note= "EMCV IRES"
FT misc_signal 2847..2849
FT /tag= e
FT /note= "Bovine alpha-lactalbumin signal peptide start
FT codon"
FT 2904..2906
FT /tag= f
FT /note= "First codon mature MN14 antibody light chain
FT gene"
FT 3543..3544
FT /tag= g
FT /note= "MN14 antibody heavy chain gene stop codon"
FT LTR 3614..4207
FT /tag= h
FT /note= "MoMuLV 3' LTR"

WO200202783-A2.
10-JAN-2002.
29-JUN-2001; 2001WO-US20714.
03-JUL-2000; 2000US-215851P.
(GALA-) GALA DESIGN INC.
Bleck GT;
WPI; 2002-154749/20.

Novel regulatory elements including nucleic acid encoding hybrid alpha-lactalbumin promoter or mutant RNA export element, for expressing one or more proteins e.g. antibodies, pharmaceutical proteins in host cells

Example 1; Fig 7; 151pp; English.

The invention relates to novel regulatory elements and vectors for the expression of one or more proteins in a host cell. The invention further provides methods of indirectly detecting the expression of a protein of interest, comprising providing the host cell transfected with a vector encoding a polycistronic sequence comprising a signal protein and a desired protein operably linked by an internal ribosome entry site (IRES), and culturing the host cell under suitable conditions so that the signal protein and the desired protein is produced, where the presence of signal protein indicates the presence of desired protein. Regulatory elements and vectors of the invention are useful for the expression of proteins of interest in a host cell. They are useful for producing an immunoglobulin (Ig), preferably secretory Ig. They are useful in the expression of one or more proteins such as erythropoietin, growth hormone, insulin, immunoglobulins, protein C, cytokines and their receptors, hormones, Von Willebrand's factor, lung surfactant, serum albumins, DNase, vascular endothelial growth factor, receptors for hormones or growth factors, rheumatoid factors, nerve growth factors, CD proteins, osteoinductive factors, immunotoxins, bone morphogenetic protein, interferons, colony stimulating factors, interleukins, superoxide dismutase, T-cell receptors, surface membrane proteins, viral antigens, transport proteins, addressins, regulatory proteins, antibodies, chimeric proteins and their fragments. The vectors are particularly useful for expressing G protein coupled receptors and other transmembrane proteins. The retroviral vectors are useful for expressing

QY 237 CCAGATCGGTCGCCGCTCAGCAGTTTCTAGA 270
 Db 270 CCAGATCGGTCGCCGCTCAGCAGTTTCTAGA 303

RESULT 12
 AAF83096
 ID AAF83096 standard; DNA; 3097 BP.

AC AAF83096;
 XX 29-JUN-2001 (first entry)
 DT Nucleotide sequence of a LTR plasmid.

XX Retrovirus; recombinase recognition sequence; RRS; LTR; recombinase;
 KW long terminal repeat; pharmaceutical; cytosolic; antinflammatory;
 KW antirheumatic; antiarthritic; antisthmatic; osteopathic; cardiac; MLV;
 KW vasotropic; neuroprotective; nontropic; cerebroprotective; antipsoriatic;
 KW antiarteriosclerotic; vulnerary; anti-HIV; antiulcer; thrombolytic;
 KW dermatological; gene therapy; ss.

XX Synthetic.

WO200125466-A1.

PD 12-APR-2001.

PF 05-OCT-2000; 2000WO-GB03837.

PR 05-OCT-1999; 99GB-0023558.

PA (OXFO-) OXFORD BIOMEDICA UK LTD.

PI Slingsby J, Kingsman SM, Rohll J, Slade A;

DR WPI; 2001-281732/29.

XX Modifying producer cells for making retrovirus by transfecting with a
 PT construct comprising a 5'- recombinase recognition sequence, long
 PT terminal repeat and 3'- recombinase recognition sequence, in presence
 PT of recombinase .

PS Example 5; Page 129-130; 133pp; English.

XX The invention relates to a method of modifying producer cells for making
 CC retrovirus by transfecting with a construct comprising a 5'- recombinase
 CC recognition sequence (RRS), long terminal repeat (LTR) and 3'- RRS, in
 CC presence of recombinase. The regulated retroviral vector produced is
 CC useful in the manufacture of a pharmaceutical composition to deliver a
 CC NOI to a target site, and in the manufacture of a medicament for
 CC diagnostic, therapeutic and/or medical applications. The recombinase
 CC assisted method is useful for introducing regulated 3'-LTR into a derived
 CC producer cell line to produce a high titer regulated retroviral vector.
 CC The vector is useful in gene therapy for treating diseases like cancers,
 CC inflammatory diseases, immunological disorders such as graft vs host
 CC diseases, autoimmune diseases such as rheumatoid arthritis, allergic
 CC diseases such as asthma, osteoporosis, cardiovascular diseases such as
 CC congestive heart failure and ischemic heart disease, neurodegenerative
 CC disorders such as multiple sclerosis, Alzheimer's disease, stroke and
 CC cerebral ischemia, atherosclerosis, thrombotic disorders, dermatological
 CC disorders such as atopic dermatitis, contact dermatitis and psoriasis,
 CC wound healing, restenosis, infectious disorders such as HIV infections,
 CC ulcers, digestive disorders such as anorexia, bulimia and cachexia, and
 CC other diseases. The present sequence represents the nucleotide sequence
 CC of a LTR plasmid.

XX Sequence 3097 BP; 770 A; 823 C; 763 G; 741 T; 0 other;

Query Match 75.1%; Score 202.8; DB 22; Length 3097;
 Best Local Similarity 95.6%; Pred. No. 9.7e-56;
 Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;

QY 1 GCTAGCTTAAGTAACGCCATTTTCAAGGCGATGGGAAAAATACATAACTGAGATAGAGA 60
 Db 1 GCTAGCTTAAGTAACGCCATTTTCAAGGCGAT -GGAAATACATAACTGAGATAGAGA 59
 QY 61 AGTTAGATCAAGGTCAGGAACAGA-GAAACAGAGAGAAATATGGCCCAACAGGATATCTG 119
 Db 60 AGTTAGATCAAGGTCAGGAACAGAGTGAACAGCTGAATATGGCCCAACAGGATATCTG 119
 QY 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGCGCAAGAACAGATTGGAACAGGAGAAT-TGGGC 177
 Db 120 TGGTAAGCAGTTCCTGCCCGCTCAGGCGCAAGAACAGATGGAACAGCTGAATATGGGC 179
 QY 178 CAAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGCGCAAGAACAGATGATGTC 236
 Db 180 CAAACAGGATATCTGTGTAAGCAGTTCCTGCCCGCTCAGGCGCAAGAACAGATGATGTC 239
 QY 237 CCAGATGGGTCGCCGCTCAGCAGTTTCTAGA 270
 Db 240 CCAGATGGGTCGCCGCTCAGCAGTTTCTAGA 273

RESULT 13

AAD28271

ID AAD28271 standard; DNA; 3671 BP.

AC AAD28271;

XX 22-APR-2002 (first entry)

DE Alpha-lactalbumin (LA) Bot vector.

XX Bovine; alpha-lactalbumin; promoter; pharmaceutical; industrial; ECMV;
 KW encephalomyocarditis virus; diagnostic; internal ribosome entry site;
 KW IRES; screening; CMV; cytomegalovirus; moloney murine leukemia virus;
 KW MoMuLV; human; pre-mRNA processing enhancer; PPE; chimeric; WPRE;
 KW woodchuck mRNA processing enhancer; botulinum; ds.

XX Chimeric - Encephalomyocarditis virus.

OS Chimeric - Bos sp.

OS Chimeric - Homo sapiens.

OS Chimeric - Moloney murine leukemia virus.

OS Chimeric - Marmota monax.

OS Chimeric - Clostridium botulinum.

OS Chimeric - Unidentified.

XX Key Location/Qualifiers

FT misc_feature 1..2053

FT /*tag= a

FT /note= "Bovine/human alpha-lactalbumin 5' flanking

FT region"

FT misc_feature 2093..2336

FT /*tag= b

FT /note= "Double mutated PPE sequence"

FT CDS 2387..2443

FT /*tag= c

FT /product= "cc49 signal peptide coding region"

FT CDS 2444..3088

FT /*tag= d

FT /product= "Bot antibody light chain Fab coding region"

FT misc_feature 3112..3686

FT /*tag= e

FT /note= "EMCV IRES"

FT CDS 3687..3745

FT /*tag= f

FT /product= "Bovine alpha-lactalbumin signal peptide

FT coding region"

FT CDS 3746..4443

FT /*tag= g

FT /product= "Bot antibody heavy chain Fab coding region"

FT misc_feature 4481..5072

FT /*tag= h

FT /note= "WPPE sequence"

FT LTR 5118..5711

Db 298 TGGTAAGCAGTTCCTGCGCCCTCAGGGCCAGGAACAGATGGAACAGCTGAATATGGGC 357
 QY 178 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAGAACAGATGGTC 236
 Db 358 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCCGCTCAGGGCCAGAACAGATGGTC 417
 QY 237 CCCAGATGCGTCCCGCCCTCAGCAGTTCCTAGTA 270
 Db 418 CCCAGATGCGTCCAGCCCTCAGCAGTTCCTAGTA 451

RESULT 10
 ID AAQ54677 standard; DNA; 636 BP.
 XX AAQ54677;
 AC
 Y 05-JUL-1994 (first entry)
 XX Promoter/Enhancer.
 XX Expression vector; promoter; enhancer; terminus; DM1; DMA2;
 KW mutant; Namalwa cell; expression plasmid; repeat terminal;
 KW Moloney retrovirus; ss.
 XX Mus musculus.
 XX JP05317077-A.
 XX 03-DEC-1993.
 XX 25-MAY-1992; 92JP-0132650.
 XX 25-MAY-1992; 92JP-0132650.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA WPI; 1994-011039/02.
 DR Prepn. of protein by recombinant animal cell - using expression
 PT vector with promoter or enhancer prep. from repeat terminal of
 PT Moloney retrovirus, and DNA for protein sequence, in Namalwa cell
 XX Claim 3; page 18; 34pp; Japanese.
 PS The sequence shows a promoter / enhancer DNA. The sequence can be
 ; recombined with a coding DNA within an expression plasmid which can
 ; be transformed into a Namalwa cell. This method can be used to prepare
 CC a useful protein efficiently.
 CC
 XX Sequence 636 BP; 185 A; 160 C; 152 G; 139 T; 0 other;
 SQ

Query Match 75.1%; Score 202.8; DB 15; Length 636;
 Best Local Similarity 95.6%; Pred. No. 4.7e-56;
 Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 QY 1 GCTAGCTTAAGTAAGCGCCATTTTGAAGCGCATGGGAAATACATACTAGAAATAGAGA 60
 Db 181 GCTAGCTTAAGTAAGCGCCATTTTGAAGCGCATGGGAAATACATACTAGAAATAGAGA 239
 QY 61 AGTTTCAGATCAAGGTTCAGGAACAGAGA-GAAACAGGAGATATGGCCAAACAGGATATCTG 119
 Db 240 AGTTTCAGATCAAGGTTCAGGAACAGAGTGAACAGCTGAATATGGCCAAACAGGATATCTG 299
 QY 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAGAACAGATTTGGAACAGGAGAAAT-TGGGC 177
 Db 300 TGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAGAACAGATTTGGAACAGGATTTGGC 359
 QY 178 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAGAACAGATTTGGC 236
 Db 360 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCCGCTCAGGGCCAGAACAGATTTGGC 419
 QY 237 CCCAGATGCGGTCCTGCGCCCTCAGCAGTTCCTAGTA 270

Db 420 CCCAGATGCGGTCCTGCGCCCTCAGCAGTTCCTAGTA 453
 RESULT 11
 ID AAT32394 standard; DNA; 702 BP.
 XX AAT32394;
 AC
 XX 28-SEP-1996 (first entry)
 XX Recombinant CMV/TAR MoMLV LTR in pMT-cat.
 DE MoMLV; long terminal repeat; CMV; HIV-1; tat protein; AIDS; retrovirus;
 KW transactivation response element; TAR; vector; antiviral; virucide;
 KW antitumour; tumour; cancer; gene therapy; ds.
 XX Chimeric Moloney murine leukemia virus;
 OS Chimeric human cytomegalovirus;
 OS Chimeric human immunodeficiency virus type 1.
 PN WO9614332-A1.
 XX 17-MAY-1996.
 PD
 XX 08-NOV-1995; 95WO-US14576.
 XX 08-NOV-1994; 94US-0336132.
 XX (CHAN/) CHANG L.
 PA Chang L;
 PI WPI; 1996-251713/25.
 DR Recombinant Moloney murine leukaemia long terminal repeat which is
 PT activated by HIV-1 Tat protein - useful in vectors for anti-viral,
 PT anti-tumour and gene therapy applications.
 PS Example 3; Page 86; 126pp; English.
 XX A recombinant MoMLV long terminal repeat (LTR) (AAT32394) was obtd.
 CC by replacing a portion of the U3 region of the Moloney murine
 CC leukaemia virus (MoMLV) LTR with the HIV-1 transactivation response
 CC element (TAR). Transfection studies showed the recombinant LTR
 CC to have a low basal activity and not to be responsive to Tat
 CC protein. This was in contrast to another recombinant LTR
 CC (AAT32382) that included the human cytomegalovirus immediate-early
 CC enhancer/promoter in addition to the TAR element.
 XX Sequence 702 BP; 170 A; 181 C; 193 G; 158 T; 0 other;
 SQ

Query Match 75.1%; Score 202.8; DB 17; Length 702;
 Best Local Similarity 95.8%; Pred. No. 4.9e-56;
 Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 QY 1 GCTAGCTTAAGTAAGCGCCATTTTGAAGCGCATGGGAAATACATACTAGAAATAGAGA 60
 Db 31 GCTAGCTTAAGTAAGCGCCATTTTGAAGCGCATGGGAAATACATACTAGAAATAGAGA 89
 QY 61 AGTTTCAGATCAAGGTTCAGGAACAGAGA-GAAACAGGAGATATGGCCAAACAGGATATCTG 119
 Db 90 AGTTTCAGATCAAGGTTCAGGAACAGAGTGAACAGCTGAATATGGCCAAACAGGATATCTG 149
 QY 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAGAACAGATTTGGAACAGGAGAAAT-TGGGC 177
 Db 150 TGGTAAGCAGTTCCTGCCCCGCTCAGGGCCAGAACAGATTTGGAACAGGATTTGGC 209
 QY 178 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGGCCAGAACAGATTTGGC 236
 Db 210 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCCGCTCAGGGCCAGAACAGATTTGGC 269

DT 05-JUN-2000 (first entry)
DE Moloney murine leukaemia virus LTR.
XX
XX Long terminal repeat; LTR; avian retrovirus;
KW constitutive transport element; CTE; replication competent;
KW host specificity; complementation; ss.
XX
XX Moloney murine leukaemia virus.
OS
XX
XX WO200006760-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17086.
XX
XX 29-JUL-1998; 98US-0094535.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Ferris AL, Hughes SH;
XX
XX WPI; 2000-205469/18.
XX
XX Novel avian retroviral vector with a modified constitutive transport
PT element allowing broadened host range is useful for gene therapy -
XX
XX Claim 2; Fig 2; 36pp; English.
XX
XX The invention relates to a novel vector comprising an avian retrovirus
CC into which a nucleotide sequence comprising a portion of a non-avian
CC retroviral LTR (long terminal repeat) has been inserted, or replaces
CC all or a portion of a native constitutive transport element (CTE). The
CC vector is replication competent in at least one non-native cell type.
CC The invention also relates to a method of delivering nucleotides into a
CC target cell via the use of the novel vector; a method of making a vector
CC that is replication competent in at least one non-native cell type,
CC comprising the excision of all or a portion of the CTE of an avian-
CC derived retrovirus and replacing the CTE with a mammalian retrovirus LTR;
CC and a chicken fibroblast cell comprising a vector derived from the avian
CC leukosis virus (ALV)-based vector RCASBP which is replication competent
CC in both mammalian and chicken cells. The vector is used as a gene therapy
CC vector, or as a research tool. By altering the CTE of the virus, the host
CC range of the retrovirus can be increased. Also, there is a reduction in
CC the need for a helper virus and a reduced chance of recombination with an
CC endogenous mammalian virus. The present sequence represents the
CC Moloney murine leukaemia virus LTR which may be inserted into an avian
CC retrovirus-derived vector which has previously had all or a portion of
CC the CTE deleted.
XX
XX Sequence 614 BP; 170 A; 164 C; 148 G; 132 T; 0 other;
XX
XX Query Match 75.1%; Score 202.8; DB 21; Length 614;
XX Best Local Similarity 95.6%; Pred. No. 4.6e-56;
XX Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
XX
XX 1 GCTAGCTTAAGTAAAGCCATTTTCCAGGCGCATGGGAAATACATACTAGATAGAGA 60
XX
XX 72 GCTAGCTTAAGTAAAGCCATTTTCCAGGCGCATGGGAAATACATACTAGATAGAGA 130
XX
XX 61 AGTTTCAGATCAAGTTCAGGACAGAGA-GAAACAGGAGATATGGGCCAAACAGGATATCTG 119
XX
XX 131 AGTTTCAGATCAAGTTCAGGACAGATGGAACAGCTGATATGGGCCAAACAGGATATCTG 190
XX
XX 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCCAAGAACAGTGGGAACAGGAGAT-TGGGC 177
XX
XX 191 TGGTAAGCAGTTCCTGCCCCGCTCAGGGCCCAAGAACAGATGGAACAGCTCAATATGGGC 250
XX
XX 178 CAACAGGATATCTGTGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCCAAGAACAGATGATGTC 236
XX
XX 251 CAACAGGATATCTGTGTAGCAGTTCCTGCCCCGCTCAGGGCCCAAGAACAGATGATGTC 310
XX
XX 237 CCCAGATGCGGTCCCGCCCTCAGCAGATTCTTAGA 270

Db 311 CCCAGATCGGTCCAGCCCTCAGCAGATTCTTAGA 344
XX
XX RESULT 9
XX AAT97160
XX ID AAT97160 standard; DNA; 635 BP.
XX
XX AC AAT97160;
XX
XX DT 11-MAY-1998 (first entry)
XX
XX DE Moloney long terminal repeat.
XX
XX KW Vaccine; B-cell malignancy; lymphoma; leukaemia; tumour;
XX gene amplification; immunotherapy; therapy; long terminal repeat;
XX LTR; MoMLV; ds.
XX
XX OS Moloney murine leukemia virus.
XX
XX PN WO9741244-A1.
XX
XX PD 06-NOV-1997.
XX
XX 25-APR-1997; 97WO-US07039.
XX
XX 06-DEC-1996; 96US-0761277.
XX
XX 01-MAY-1996; 96US-0644664.
XX
XX (GENI-) GENITOPE CORP.
XX
XX Denney DW;
XX
XX WPI; 1997-549743/50.
XX
XX Multivalent vaccine to treat B cell lymphoma or leukaemia -
XX comprises at least 2 different recombinant variable regions of
XX immunoglobulin molecules derived from B cell lymphoma cells
XX
XX Example 1; Page 104-105; 177pp; English.
XX
XX This DNA sequence comprises the Moloney murine leukaemia virus
XX (MoMLV) long terminal repeat (LTR) obtained from a plasmid
XX containing MoMLV DNA by restriction enzyme digestion. The LTR is
XX utilised in expression vectors designed for efficient expression of
XX genes in eukaryotic cells. The invention provides vectors and
XX improved methods for the expression and co-amplification of genes
XX encoding recombinant proteins in cultured cells. The methods
XX permit the isolation of cell lines which have co-amplified input
XX recombinant sequences which encode an amplifiable marker, one or
XX more expression vectors encoding a protein of interest and
XX optionally a selectable marker. The amplified cells provide large
XX quantities of recombinant proteins suitable for immunotherapy for
XX treatment of lymphomas and leukaemias. The methods permit the
XX production of custom vaccines, including multivalent vaccines that
XX reflect the degree of somatic variation found in a patient's tumour.
XX
XX Sequence 635 BP; 184 A; 159 C; 153 G; 139 T; 0 other;
XX
XX Query Match 75.1%; Score 202.8; DB 18; Length 635;
XX Best Local Similarity 95.6%; Pred. No. 4.7e-56;
XX Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
XX
XX 1 GCTAGCTTAAGTAAAGCCATTTTCCAGGCGCATGGGAAATACATACTAGATAGAGA 60
XX
XX 179 GCTAGCTTAAGTAAAGCCATTTTCCAGGCGCATGGGAAATACATACTAGATAGAGA 237
XX
XX 61 AGTTTCAGATCAAGTTCAGGACAGAGA-GAAACAGGAGATATGGGCCAAACAGGATATCTG 119
XX
XX 238 AGTTTCAGATCAAGTTCAGGACAGATGGAACAGCTGATATGGGCCAAACAGGATATCTG 297
XX
XX 120 TGGTAAGCAGTTCCTGCCCC-GCTCAGGGCCCAAGAACAGTGGGAACAGGAGAT-TGGGC 177

PI Von Laer M;
 XX WPI; 2001-367622/38.
 DR P-PSDB; AAB86198.
 XX
 PT New nucleic acid encoding membrane-anchored gp41 fusion protein, useful
 PT for gene therapy of human immunodeficiency virus (HIV) infection,
 PT prevents entry of virus into cells
 XX
 XX Claim 11; Page 28-30; 39pp; German.
 PS
 CC This invention describes a novel nucleic acid (I) comprising elements
 CC that encode a signal peptide (SP) that provides transfer of expressed
 CC polypeptide into the endoplasmic reticulum, a fragment (FI) of HIV gp41
 CC protein, containing a segment from a heptad repeat region, a
 CC transmembrane anchor (MSD) of a type I membrane protein and a flexible
 CC linker (hinge) linking FI and MSD, therefore the formula of (I) is
 CC SP-FI-hinge-MSD. The products of the invention have antiviral activity
 CC and act as HIV replication inhibitors. Vectors containing (I), also T
 CC lymphocytes or hematopoietic stem cells transfected in vitro with (I),
 CC are used in gene therapy of HIV infection.
 XX
 SQ Sequence 4148 BP; 880 A; 1177 C; 1103 G; 988 T; 0 other;
 Query Match 77.1%; Score 208.2; DB 22; Length 4148;
 Best Local Similarity 97.1%; Pred. No. 1.9e-57;
 Matches 265; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
 QY 1 GCTAGCTTAAGTACGCGCATTTTGCAGGCGCATGGGAAAAATACATACTGAGAATAGAGA 60
 DB 3586 GCTAGCTTAAGTAA-GCATTTTGCAGGCGCAT-GGAAAAATACATACTGAGAATAGAGA 3643
 QY 61 AGTTCAGATCAAGTTCAGGACAGAGAACAGGAGAAATATGGCCAAACAGAGATATCTGT 120
 DB 3644 AGTTCAGATCAAGTTCAGGACAGAGAACAGGAGAAATATGGCCAAACAGAGATATCTGT 3703
 QY 121 GGTAAGCAGTTCCTGCCCC-CCTCAGGCGCCAGACAGTTCGAAACAGAGAGAT-TGGGCC 178
 DB 3704 GGTAAGCAGTTCCTGCCCCCGGCTCAGGCGCCAGACAGTTCGAAACAGAGAGAT-TGGGCC 3763
 QY 179 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-CGTCAGGCGCCAAACAGAGATGGTCC 237
 DB 3764 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCCCGGCTCAGGCGCCAAACAGAGATGGTCC 3823
 QY 238 CCAGATGCGGTCCCGCCCTCAGCAGTTCCTAGTA 270
 O 3824 CCAGATGCGGTCCCGCCCTCAGCAGTTCCTAGTA 3856
 RESULT 7
 AAF83043
 ID AAF83043 standard; DNA; 508 BP.
 XX
 AC AAF83043;
 XX
 DT 29-JUN-2001 (first entry)
 DE
 XX PCR amplified MLV U3 region from pHIT11.
 XX
 KW Retrovirus; recombinase recognition sequence; RRS; LTR; recombinase;
 KW long terminal repeat; pharmaceutical; cytostatic; antiinflammatory;
 KW antirheumatic; antithyroid; antidiabetic; osteopathic; cardiant; MLV;
 KW vasotrophic; neuroprotective; nootropic; cerebroprotective; antipsoriatic;
 KW antiarteriosclerotic; vulnerary; anti-HIV; antiulcer; thrombolytic;
 KW dermatological; gene therapy; ss.
 XX
 OS Murine leukemia virus.
 XX
 PN WO200125466-A1.
 XX
 PD 12-APR-2001.
 XX
 PF 05-OCT-2000; 2000WO-GB03837.

XX
 PR 05-OCT-1999; 99GB-0023558.
 XX
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 PI Slingsby J, Kingsman SM, Rohll J, Slade A;
 DR WPI; 2001-281732/29.
 XX
 PT Modifying producer cells for making retrovirus by transfecting with a
 PT construct comprising a 5'- recombinase recognition sequence, long
 PT terminal repeat and 3'- recombinase recognition sequence, in presence
 PT of recombinase
 XX
 PS Example 7; Page 51; 133pp; English.
 XX
 CC The invention relates to a method of modifying producer cells for making
 CC retrovirus by transfecting with a construct comprising a 5'- recombinase
 CC recognition sequence (RRS), long terminal repeat (LTR) and 3'- RRS, in
 CC presence of recombinase. The regulated retroviral vector produced is
 CC useful in the manufacture of a pharmaceutical composition to deliver a
 CC NOI to a target site, and in the manufacture of a medicament for
 CC diagnostic, therapeutic and/or medical applications. The recombinase
 CC assisted method is useful for introducing regulated 3'-LTR into a derived
 CC producer cell line to produce a high titer regulated retroviral vector.
 CC The vector is useful in gene therapy for treating diseases like cancers,
 CC inflammatory diseases, immunological disorders such as graft vs host
 CC disease, autoimmune diseases such as rheumatoid arthritis, allergic
 CC diseases such as asthma, osteoporosis, cardiovascular diseases such as
 CC congestive heart failure and ischemic heart disease, neurodegenerative
 CC disorders such as multiple sclerosis, Alzheimer's disease, stroke and
 CC cerebral ischemia, atherosclerosis, thrombotic disorders, dermatological
 CC disorders such as atopic dermatitis, contact dermatitis and psoriasis,
 CC wound healing, restenosis, infectious disorders such as HIV infections,
 CC ulcers, digestive disorders such as anorexia, bulimia and cachexia, and
 CC other diseases. The present sequence represents a PCR amplified murine
 CC leukemia virus (MLV) U3 region from pHIT11, used to replace Equine
 CC infectious anemia virus (EIAV) PPT/U3 sequence. This is used in the
 CC construction of EIAV vectors with LTR driven open reading frames.
 XX
 SQ Sequence 508 BP; 135 A; 129 C; 129 G; 115 T; 0 other;
 Query Match 75.1%; Score 202.8; DB 22; Length 508;
 Best Local Similarity 95.6%; Pred. No. 4.2e-56;
 Matches 262; Conservative 0; Mismatches 7; Indels 5; Gaps 5;
 QY 1 GCTAGCTTAAGTAAAGCCCATTTTGCAGGCGCATGGGAAAAATACATACTGAGAATAGAGA 60
 DB 58 GCTAGCTTAAGTAAAGCCCATTTTGCAGGCGCAT-GGAAAAATACATACTGAGAATAGAGA 116
 QY 61 AGTTCAGATCAAGGTCAGGACAGAGA-GAAACAGGAGAAATATGGCCAAACAGAGATATCTG 119
 DB 117 AGTTCAGATCAAGGTCAGGACAGAGATGGAAACAGCTGAATATGGCCAAACAGAGATATCTG 176
 QY 120 TGTAGCAGAGTTCCTGCCCC-CGTCAGGCGCCAGACAGTTCGAAACAGAGAGAT-TGGGCC 177
 DB 177 TGTAGCAGAGTTCCTGCCCCCGGCTCAGGCGCCAGACAGATGGAAACAGCTGAATATGGCC 236
 QY 178 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-CGTCAGGCGCCAAACAGAGATGGTC 236
 DB 237 CAACAGGATATCTGTGTAAGCAGTTCCTGCCCCCGGCTCAGGCGCCAAACAGAGATGGTC 296
 QY 237 CCAGATGCGGTCCCGCCCTCAGCAGTTCCTAGTA 270
 DB 297 CCAGATGCGGTCCAGCCCTCAGCAGTTCCTAGTA 330
 RESULT 8
 AA292704
 ID AA292704 standard; cDNA; 614 BP.
 XX
 AC AA292704;
 XX

XX The invention relates to a viral vector system for preparing recombinant
 CC adeno-associated virus (AAV) particles comprising: at least two plasmid
 CC vectors (ABL58983 and ABL58984) that include the two inverted terminal
 CC repeats (ITR) of AAV and additional sequences and plasmid vectors without
 CC ITRs but containing the rep and cap genes of AAV required for replication
 CC and packaging. The system is useful for producing recombinant AAV for
 CC production of a wide range of therapeutic glycoproteins in eukaryotic
 CC cells. The system provides efficient, large scale production of
 CC heterologous proteins in mammalian cells, without requiring an adenovirus
 CC helper. It is not toxic to host cells and does not cause lysis, so
 CC produced proteins are highly pure. The present sequence is that of the
 CC pAIM E1B5K plasmid of the invention.
 XX
 SQ Sequence 6575 BP; 1623 A; 1566 C; 1695 G; 1691 T; 0 other;

Query Match 91.1%; Score 246; DB 24; Length 6575;
 Best Local Similarity 99.3%; Pred. No. 1.1e-69;
 Matches 268; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1 GCTAGCTTAAGTAAAGCCATTTTGAAGGCATGGGAAATACATAACTCAGAAATAGAGA 60
 432 GCTAGCTTAAGTAAAGCCCA-TTTCAGGCAAT-GGAAATACATAACTCAGAAATAGAGA 489
 61 AGTTCAGATCAAGGTTCAGGACAGAGAAACAGGAGAAATATGGGCCAAACAGGATATCTGT 120
 490 AGTTCAGATCAAGGTTCAGGACAGAGAAACAGGAGAAATATGGGCCAAACAGGATATCTGT 549
 121 GGTAAAGCATCTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAGAAATGGGCCAA 180
 550 GGTAAAGCATCTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAGAAATGGGCCAA 609
 181 ACAGGATATCTGTGTAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAG 240
 610 ACAGGATATCTGTGTAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAG 669
 241 GATCGGTGCTCCCGCTCAGCAGTTCTCTAG 270
 670 GATCGGTGCTCCCGCTCAGCAGTTCTCTAG 699

RESULT 5
 AAA96220
 ID AAA96220 standard; DNA; 9830 BP.
 XX
 AC AAA96220;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Nucleotide sequence of pTRONIN, a retroviral vector.
 XX
 Retroviral vector; inflammatory disorder; dermatological disorder;
 cardiovascular disorder; autoimmune disease; neurological disorder;
 cancer; gene therapy; ss.
 Synthetic.
 WO200056910-A1.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-GB01091.
 XX
 PR 22-MAR-1999; 99GB-0006615.
 XX
 PA (OXFO-) OXFORD BIOMEDICA UK LTD.
 XX
 PI Uden M, Mitrophanous K;
 XX
 DR WPI; 2000-628271/60.
 XX
 PT Retroviral vector for delivering one or more nucleotides of interest to
 a target site has functional and non-functional splice donor and splice

PT acceptor sites -
 XX
 PS Example 8; Fig 33; 148pp; English.
 XX
 CC The specification describes a retroviral vector which comprises a
 CC nucleotide sequence of interest flanked by functional splice donor
 CC sites and functional splice acceptor sites. The vector is derived from
 CC a retroviral pro-vector. The retroviral vector is useful for preparing
 CC pharmaceutical compositions to deliver one or more nucleotide sequences
 CC of interest to a target sit. The retroviral vectors are especially
 CC useful for treating inflammatory disorders, cancers, dermatological
 CC disorders, cardiovascular disorders, autoimmune diseases and neurological
 CC disorders. The retroviral vector is useful in gene therapy. The present
 CC sequence represents the retroviral vector pTRONIN, a vector of the
 CC invention.
 XX

SQ Sequence 9830 BP; 2209 A; 2659 C; 2688 G; 2274 T; 0 other;
 Query Match 79.7%; Score 215.2; DB 21; Length 9830;
 Best Local Similarity 94.1%; Pred. No. 1.5e-59;
 Matches 256; Conservative 0; Mismatches 13; Indels 3; Gaps 3;
 QY 1 GCTAGCTTAAGTAAAGCCATTTTGAAGGCATGGGAAATACATAACTCAGAAATAGAGA 60
 6141 GCTAGCTTAAGTAAAGCCATTTTGAAGGCAT-GGAAATACATAACTCAGAAATAGAGA 6199
 61 AGTTCAGATCAAGGTTCAGGACAGAGAAACAGGAGAAATATGGGCCAAACAGGATATCTGT 119
 6200 AGTTCAGATCAAGGTTCAGGACAGAGAAACAGGAGAAATATGGGCCAAACAGGATATCTGT 6259
 120 TGGTAAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAGAAATGGGCC 178
 6260 TGGTAAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGACAGAGAAATGGGCC 6319
 179 AAACAGGATATCTGTGTAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGTCCC 238
 6320 AAACAGGATATCTGTGTAAGCAGTTCTGCTCCCGCTCAGGGGCAAGAACAGAGTTGGTCCC 6379
 239 CAGATGCGGTGCTCCCGCTCAGCAGTTTCTAGA 270
 6380 CAGATGCGGTGCTCCCGCTCAGCAGTTTCTAGA 6411

RESULT 6
 AAH20890
 ID AAH20890 standard; DNA; 4148 BP.
 XX
 AC AAH20890;
 XX
 DT 24-AUG-2001 (first entry)
 XX
 DE Vector containing HIV gp41 DNA SEQ ID 1.
 XX
 KW Transmembrane anchor; gene therapy; endoplasmic reticulum; gp41;
 KW antiviral; HIV replication inhibitor; T lymphocyte; viral infection;
 KW hematopoietic stem cell; ds.
 XX
 OS Human immunodeficiency virus type 1.
 XX
 FH Key Location/Qualifiers
 CDS 1438..1773
 FT /*tag= a
 FT
 XX
 PN WO200137881-A2.
 XX
 PD 31-MAY-2001.
 XX
 PF 24-NOV-2000; 2000WO-EPI1733.
 XX
 PR 25-NOV-1999; 99DE-1057838.
 XX
 PA (PETT-) PETTE INST HEINRICH.
 XX

This sequence encodes an epitope-tagged TATA-box binding protein (TBP) that is expressed by the transgenic non-human animals of the invention. The animals are used to produce TBP. TBP is used to isolate and

PT mammalian gene expression, is based on the interferon regulating
 XX factor-1 and its binding site
 PS
 XX Claim 1; Page 9; 19pp; English.

CC The present invention describes a promoter-transactivator system for
 CC inducible high-level expression of mammalian genes, and optionally
 CC for control of cell growth. The promoter-transactivator system
 CC comprises: (i) a promoter construct (IRFE promoter); and (ii)
 CC transactivator construct encoding a fusion protein of IRF-1 (interferon
 CC regulating factor-1) and the estrogen receptor (ER). The IRFE promoter
 CC construct has the structure (MPSV-E)-(IRF-1 binding site)-(CMV)-DNA
 CC where MPSV-E indicates the myeloproliferative sarcoma virus enhancer
 CC repeat given in AAC64591; IRF-1 binding site given in AAC64592; and CMV
 CC is the cytomegalovirus minimal promoter given in AAC64593 or their
 CC functionally equivalent variants with one or more nucleotides
 CC substituted, inserted or deleted. The promoter transactivator system is
 CC a transcription regulator. Increased transcription results from binding
 CC of the IRFE promoter to the transactivator, which is activated by
 CC oestradiol or other ER ligands (these compounds displace the heat-shock
 CC protein 90 which normally binds to the IRF-1/ER fusion, preventing its
 CC activation). The system is used for increasing production of
 CC therapeutically active proteins and where a IRF-1-green fluorescent
 CC protein-human ER construct is used, for fluorescence-activated cell
 CC sorting/analysis of transformed cells.

XX Sequence 270 BP; 85 A; 61 C; 74 G; 50 T; 0 other;

Query Match 100.0%; Score 270; DB 21; Length 270;
 Best Local Similarity 100.0%; Pred. No. 3.7e-78;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTAGCTTTAAGTAAGCCATTTTTCGAAGCATGGGAAATACATACTGAGATAGAGA 60
 Db 1 GCTAGCTTTAAGTAAGCCATTTTTCGAAGCATGGGAAATACATACTGAGATAGAGA 60

QY 61 AGTTCAGATCAAGTTCAGAACAGAGAAACAGAGAAATATGGGCCAAACAGGATATCTGT 120
 Db 61 AGTTCAGATCAAGTTCAGAACAGAGAAACAGAGAAATATGGGCCAAACAGGATATCTGT 120

QY 121 GGTAAAGCATCTCTCCCGCTCAGGCGCAAGAACAGAGTTGGAAACAGAGAAATGGGCCAA 180
 Db 121 GGTAAAGCATCTCTCCCGCTCAGGCGCAAGAACAGAGTTGGAAACAGAGAAATGGGCCAA 180

QY 181 ACAGGATATCTGTGTAGACAGTTCCTGCCCGCTCAGGCGCAAGAACAGAGATGTCGCCCA 240
 Db 181 ACAGGATATCTGTGTAGACAGTTCCTGCCCGCTCAGGCGCAAGAACAGAGATGTCGCCCA 240

QY 241 GATCGGTCCTCCCGCTCAGCAGTTTCTAGA 270
 Db 241 GATCGGTCCTCCCGCTCAGCAGTTTCTAGA 270

RESULT 2
 AAV18096
 ID AAV18096 standard; DNA; 8298 BP.
 XX
 AC AAV18096;
 XX
 DT 04-AUG-1998 (first entry)
 XX
 DE pMCLDHPA tricitronic vector for the expression of hmAb45-TNF alpha.
 XX
 KW Circular; antibody-cytokine fusion protein; tricitronic vector;
 KW TNF alpha; IL-2; IRES; internal ribosome entry site; ds.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT promoter 1..904
 FT /tag= a
 FT /note= "CMV promoter with an upstream MPSV enhancer"
 FT 905..976
 FT Intron

FT /tag= b
 FT /number= Intron 1
 FT 977..1018
 FT /tag= c
 FT /note= "Partial leader sequence"
 FT 1019..1106
 FT /tag= d
 FT /note= "5'UTR from poliovirus"
 FT 1107..1433
 FT /tag= e
 FT /product= "Light chain hmAb425, variable region"
 FT 1107..1115
 FT /tag= f
 FT /note= "Rest of the leader sequence"
 FT 1434..1595
 FT /tag= g
 FT /number= Intron 2
 FT 1596..1913
 FT /tag= h
 FT /product= "Light chain hmAb425, constant region"
 FT 1914..2028
 FT /tag= i
 FT /note= "5' UTR from poliovirus"
 FT 2029..2159
 FT /tag= j
 FT /note= "Poliovirus derived internal ribosome entry
 FT site"
 FT 1260..2581
 FT /tag= Intron 3
 FT 2582..4537
 FT /tag= l
 FT /note= "Heavy chain hmAb425 fused to TNF alpha"
 FT 4565..5279
 FT /tag= m
 FT /note= "Comprises of a 5' UTR from poliovirus, an
 FT internal ribosome entry site and Intron 4"
 FT 5280..5876
 FT /tag= n
 FT /product= "Puromycin acetyl transferase"
 FT /note= "Selection marker"
 FT 5929..6181
 FT /tag= o
 FT /standard_name= SV40 PolyA
 XX
 PN WO9811241-A1.
 XX
 PD 19-MAR-1998.
 XX
 PF 02-SEP-1997; 97WO-EP04765.
 XX
 PR 30-SEP-1996; 96EP-0115635.
 PR 16-SEP-1996; 96EP-0114820.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 XX Bruemmer W, Burge C, Dunker R, Hauser H, Mielke C;
 PI Rieke E, Von Hoegen I, Welge T;
 XX
 DR WPI: 1998-207400/18.
 DR P-PSDB; AAW48647, AAW48648, AAW48649, AAW48650, AAW48651.
 XX
 PT Oligo:citrionic expression vector - useful for production of, e.g.
 PT MAb425/TNF-a or MAb425/IL-2 antibody fusion protein
 XX
 PS Claim 11; Fig 15; 89pp; English.
 XX
 CC The present sequence represents a new pMCLDHPA tricitronic vector for
 CC the expression of an antibody-cytokine fusion protein, hmAb425-TNF
 CC alpha. hmAb425-TNF alpha comprises of the TNF alpha fused to the
 CC C-terminus of the heavy chain of humanized monoclonal antibody 425.
 CC The TNF alpha sequence can be substituted by the IL-2 sequence. The
 CC hmAb425 has specificity for the human EGF receptor. The vector also
 CC contains a strong promoter/enhancer unit, a selection marker gene and

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2003, 22:29:12 ; Search time 263.478 Seconds
(without alignments)
2307.741 Million cell updates/sec

Title: us-09-980-548-1
Perfect score: 270
Sequence: 1 gtagcttaagtaacgcat.....cgccctcagcagtttctaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

-searched: 2185239 segs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	270	100.0	270	21 AAC64591	Myeloproliferative
2	270	100.0	8298	19 AAV18096	PMCLDHPA tricitro
3	254.6	94.3	3263	20 AAV08560	Transgene for epit
4	246	91.1	6575	24 ABL58984	AAV expression vec
5	215.2	79.7	9830	21 AAA96220	Nucleotide sequenc
6	208.2	77.1	4148	22 AAH20890	Vector containing
7	202.8	75.1	508	22 AAF83043	PCR amplified MLV
8	202.8	75.1	614	21 AA292704	Moloney murine leu
9	202.8	75.1	635	18 AAT97160	Moloney long termi

10	202.8	75.1	636	15 AAQ54677	Promoter/Enhancer.
11	202.8	75.1	702	17 AAT32394	Recombinant CMV/TA
12	202.8	75.1	3097	22 AAF83096	Nucleotide sequenc
13	202.8	75.1	3671	24 AAD28271	Alpha-lactalbumin
14	202.8	75.1	4207	24 AAD28267	Cytomegalovirus (C
15	202.8	75.1	4207	24 AAD28306	Cytomegalovirus (C
16	202.8	75.1	4210	24 AAD28268	Cytomegalovirus (C
17	202.8	75.1	4210	24 AAD28307	Cytomegalovirus (C
18	202.8	75.1	4661	24 AAD28273	Alpha-lactalbumin
19	202.8	75.1	4661	24 AAD28312	Alpha-lactalbumin
20	202.8	75.1	4776	20 AAX77617	Expression constru
21	202.8	75.1	4776	20 AAX77614	Expression constru
22	202.8	75.1	4924	21 AAZ34937	Retrovirus vector.
23	202.8	75.1	4924	22 AAF30946	Retrovirus vector.
24	202.8	75.1	5109	18 AAT76800	PLJ Rev m10 retrov
25	202.8	75.1	5130	24 AAD28311	LSRNL vector. Chl
26	202.8	75.1	5176	18 AAT76801	PLJ mutant Rev m10
27	202.8	75.1	5177	20 AAZ05997	Bovine scavenger r
28	202.8	75.1	5594	19 AAV33629	GENSA 981, a monom
29	202.8	75.1	5617	24 AAD32077	Human alpha-1-anti
30	202.8	75.1	5689	20 AAZ11445	Retroviral vector
31	202.8	75.1	5689	20 AAX61061	Retroviral vector
32	202.8	75.1	5689	21 AAA96208	Nucleotide sequenc
33	202.8	75.1	5691	24 AAD28274	Alpha-lactalbumin
34	202.8	75.1	5691	24 AAD28313	Alpha-lactalbumin
35	202.8	75.1	5711	24 AAD28310	Alpha-lactalbumin
36	202.8	75.1	5715	21 AAZ34936	Retrovirus vector.
37	202.8	75.1	5715	22 AAF30945	Vector used in inv
38	202.8	75.1	5731	24 AAD32078	Human alpha-1-anti
39	202.8	75.1	5732	24 AAD28269	Mouse mammary tumo
40	202.8	75.1	5732	24 AAD28308	Mouse mammary tumo
41	202.8	75.1	5865	19 AAV04002	Retroviral vector
42	202.8	75.1	5874	20 AAX90484	Plasmid retroviral
43	202.8	75.1	6026	24 AAD32075	Human albumin prom
44	202.8	75.1	6140	24 AAD32076	Human albumin prom
45	202.8	75.1	6141	20 AAX90481	Plasmid retroviral

ALIGNMENTS

RESULT 1
AAC64591
ID AAC64591 standard; DNA; 270 BP.

XX AAC64591;

XX 15-FEB-2001 (first entry)

DT Myeloproliferative sarcoma virus enhancer sequence SEQ ID NO:1.

DE Myeloproliferative sarcoma virus; MPSV; MPSV-E; enhancer; CMV;

XX Cytomegalovirus; firefly; IRF-1 binding site; minimal promoter;

KW interferon regulatory factor 1 binding site; transcription regulator;

KW promoter transactivator system; ds.

XX Myeloproliferative sarcoma virus.

OS EPI046710-A1.

PN 25-OCT-2000.

PD 23-APR-1999; 99EP-0108068.

XX 23-APR-1999; 99EP-0108068.

XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA Mueller PP, Geserick C, Schroeder K, Hauser H;

PI WPI; 2000-648930/63.

DR Promoter-transactivator system, useful for inducing high level

PT

```
protein_bind      /bound_moiety="glucocorticoid receptor"
enhancer          263..275
repeat_region     /bound_moiety="nuclear factor (NF)"
protein_bind      287..294
protein_bind      /rpt_type=direct
protein_bind      320..392
protein_bind      321..326
protein_bind      /bound_moiety="glucocorticoid receptor"
protein_bind      338..349
enhancer          /bound_moiety="nuclear factor (NF)"
protein_bind      361..368
protein_bind      396..400
protein_bind      /bound_moiety="glucocorticoid receptor"
CAAT_signal       508..512
TATA_signal       559..565
mRNA              590..733
repeat_region     /partial
repeat_region     590..657
repeat_region     /note="R region"
repeat_region     658..733
repeat_region     /note="US region"
repeat_region     721..733
repeat_region     /rpt_type=inverted
BASE COUNT       195 a 193 c 176 g 168 t 1 others
ORIGIN
```

```
Query Match      76.4%; Score 206.4; DB 14; Length 733;
Best Local Similarity 94.1%; Pred. No. 1.6e-53;
Matches 257; Conservative 0; Mismatches 12; Indels 4; Gaps 4;

Qy 1 GCTAGCTTAAGTAACGCCATTTTGCAGGCGATGGGAAATACATACTAGAGATAGAGA 60
Db 170 GCTAGCTTAAGTAACGCCATTTTGCAGGCGATGGGAAATACATACTAGAGATAGAGA 228
Qy 61 AGTTGAGATCAAGTCAGGACAGAG- AAACAGGAGAAATATGGCCAAACAGGATATCTG 119
Db 229 AGTTGAGATCAAGTCAGGACAGAGAAACAGCTGAAATATGGCCAAACAGGATATCTG 288
Qy 120 TGGTAAGCAGTTCTTGGCCCT- GCTCAGGGCCAAAGAACAGTTGGAACAGGAAATTTGGGCC 178
Db 289 TGGCAAGCAGTTCTTGGCCCTCAGGGCCAAAGAACAGATGAATCAGCTGAATTTGGGCC 348
Qy 179 AAACAGGATATCTGTTAGCAGTTCTTGGCCCT- GCTCAGGGCCAAAGAACAGATGTTGCC 237
Db 349 AAACAGGATATCTGTTAGCAGTTCTTGGCCCTCAGGGCCAAAGAACAGATGTTGCC 408
.y 238 CCAGATGCGGTCCTCGCCCTCAGCAGTTTCTAGA 270
Db 409 CCAGATGCGGTCCTCGCCCTCAGCAGTTTCTAGA 441
```

Search completed: March 5, 2003, 23:28:17
Job time : 1503.61 secs

SignalPEPTIDEkodiert
T-20-Fusionsprotein mit Membrananker
/codon_start=1
/transl_table=11
/protein_id="CAC42734.1"
/db_xref="GI:14546735"
/translation="MGAGATGRAMDGPRLLLLLLVSLGARSYTLIHLISLBESON
QOEKNEQELLELDKWSLWNFRSVPDCCKPCICTLPIVYCSILAADVVLVAYIA
FKRWNRGIL"
1438..1527
1438..1773
/note="STHM (Vektor-Insert, das den fuer das
Membran-verankerte T-20-PEPTIDEkodierenden Bereich
enthaelt)"
1528..1635
/note="Fuer T-20 kodierender Bereich"
1636..1683
/note="Fuer Hinge kodierender Bereich"
1684..1773
/note="Fuer Membrananker kodierender Bereich"
1775..1780
/note="Sali-Schnittstelle"
1787..1792
/note="Sphi-Schnittstelle"
1793..2421
/note="Polio-IRES"
1793..1798
/note="HindIII-Schnittstelle"
2380..3267
/note="neor CDNA"
2422..2429
/note="NotI-Schnittstelle"
3002..3007
/note="Sphi-Schnittstelle"
3447..3452
/note="HindIII-Schnittstelle"
3586..3591
/note="NheI-Schnittstelle"
4000..4007
/note="AscI-Schnittstelle"
4030..4035
/note="XmaI-Schnittstelle"
BASE COUNT 880 a 1177 c 1103 g 988 t
ORIGIN

Query Match 77.1%; Score 208.2; DB 6; Length 4148;
Best Local Similarity 97.1%; Pred. No. 5.5e-54;
Matches 265; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
Qy 1 GCTAGCTTAAGTAACGCCATTTTGCAGGCGATGGGAAAAATACATAACTGAGAATAGAGA 60
|||||
3586 GCTAGCTTAAGTAA-GCCATTTTGCAGGCGAT-GGAAAAATACATAACTGAGAATAGAGA 3643
Qy 61 AGTTTCAGATCAGGTCAGGAACAGAGAGAGATATGGGCCAACAGGATATCTGT 120
|||||
3644 AGTTTCAGATCAGGTCAGGAACAGAGAGAGATATGGGCCAACAGGATATCTGT 3703
Qy 121 GCTAGCAGCTTCTGCCCC-GCTCAGGCGCAAGAACAGTGGACAGAGGAAT-TGGGCC 178
|||||
3704 GCTAGCAGCTTCTGCCCCGGCTCAGGCGCAAGAACAGTGGACAGAGGAATATGGGCC 3763
Qy 179 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGCGCAAGAACAGATGGTCC 237
|||||
3764 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCCGGCTCAGGCGCAAGAACAGATGGTCC 3823
Qy 238 CCAGATGGCGTCCGCCCTCAGCAGTTTCTAGA 270
|||||
3824 CCAGATGGCGTCCGCCCTCAGCAGTTTCTAGA 3856
Db

RESULT 14
AR028672
LOCUS

AR028672 9318 bp DNA linear PAT 29-SEP-1999

DEFINITION Sequence 6 from patent US 5858744.
ACCESSION AR028672
VERSION AR028672.1 GI:5940645
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 9318)
AUTHORS Baum,C., Stocking-Harbers,C. and Ostertag,W.
TITLE Retroviral vector hybrids and the use thereof for gene transfer
JOURNAL Patent: US 5858744-A 6 12-JAN-1999;
FEATURES Location/Qualifiers
source 1..9318
/organism="unknown"

BASE COUNT 2377 a 2217 c 2364 g 2358 t
ORIGIN

Query Match 77.1%; Score 208.2; DB 6; Length 9318;
Best Local Similarity 97.1%; Pred. No. 6e-54;
Matches 265; Conservative 0; Mismatches 3; Indels 5; Gaps 5;
Qy 1 GCTAGCTTAAGTAACGCCATTTTGCAGGCGATGGGAAAAATACATAACTGAGAATAGAGA 60
|||||
Db 5899 GCTAGCTTAAGTAA-GCCATTTTGCAGGCGAT-GGAAAAATACATAACTGAGAATAGAGA 5956
Qy 61 AGTTTCAGATCAGGTCAGGAACAGAGAGAGATATGGGCCAACAGGATATCTGT 120
|||||
Db 5957 AGTTTCAGATCAGGTCAGGAACAGAGAGAGATATGGGCCAACAGGATATCTGT 6016
Qy 121 GCTAGCAGTTCCTGCCCC-GCTCAGGCGCAAGAACAGTGGACAGAGGAAT-TGGGCC 178
|||||
Db 6017 GCTAGCAGTTCCTGCCCCGGCTCAGGCGCAAGAACAGTGGACAGAGGAATATGGGCC 6076
Qy 179 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCC-GCTCAGGCGCAAGAACAGATGGTCC 237
|||||
Db 6077 AAACAGGATATCTGTGTAAGCAGTTCCTGCCCCGGCTCAGGCGCAAGAACAGATGGTCC 6136
Qy 238 CCAGATGGCGTCCGCCCTCAGCAGTTTCTAGA 270
|||||
Db 6137 CCAGATGGCGTCCGCCCTCAGCAGTTTCTAGA 6169

RESULT 15
MLM3LTR

LOCUS MLM3LTR 733 bp ss-RNA linear VRL 09-DEC-1994
DEFINITION Moloney murine leukemia virus TB 3' LTR region.
ACCESSION M24204
VERSION M24204.1 GI:600759
KEYWORDS LTR.
SOURCE Moloney murine leukemia virus (strain TB) RNA.
ORGANISM Moloney murine leukemia virus
VIRUSES; Retroviral viruses; Retroviridae; Gammaretrovirus.

REFERENCE 1 (bases 1 to 733)
AUTHORS Yuen,P.H. and Szurek,P.F.
TITLE The reduced virulence of the thymotropic Moloney murine leukemia
virus derivative MoMuLV-TB is mapped to 11 mutations within the U3
region of the long terminal repeat
J. Virol. 63 (2), 471-480 (1989)

JOURNAL 89094971
MEDLINE 2783465
PUBMED

COMMENT On Dec 10, 1994 this sequence version replaced gi:341122.
FEATURES Location/Qualifiers
source 1..733
/organism="Moloney murine leukemia virus"
/strain="TB"
/db_xref="taxon:11801"

repeat_region 140..590

repeat_region 140..153

repeat_region 245..319

protein_bind 246..251

FEATURES
source Location/Qualifiers
1. .2854
/organism="Myeloproliferative sarcoma virus"
/db_xref="taxon:11813"
973. 2001
/note="v-mos protein"
/codon_start=1
/protein_id="AAA46580.1"
/db_xref="GI:332209"
/translation="MPSPLSLCRVLPRLSPSVDSRCSIPLVAPRKAGKFLGTTPP
RAGLPRLAFLFSDVQCLMHLRIGSGGVSFKATYHGVPAIKOVNCKPKDLRAS
QPSFAELNIALRHDIIVVAATSTPPEDSNLGTIIIEFGNVTLHOVIYGATRS
PEPLSCRELKCLKYSLDVNGLLFLHSQSIHLKPKPANILISEQDVKISDFG
CSQQLDRLCRSPHHIGGTYTHQAPILKGEIATPKADIYSFGILWMTREMPYS
GEPOYVOYAVVAYNLRLPSLTGAVTASLTGTLQNIIONCWEARLQRPGEALLORDL
KAFRGALC"
LTR 2268. 2854
/note="3' long terminal repeat"
repeat_region 2383. .2456
/note="tandem repeat A"
repeat_region 2457. .2530
/note="tandem repeat B"
BASE COUNT 696 a 814 c 709 g 635 t
ORIGIN 1 bp upstream of HindIII site.
Query Match 82.6%; Score 223; DB 14; Length 2854;
Best Local Similarity 98.5%; Pred. No. 1.3e-58;
Matches 267; Conservative 0; Mismatches 0; Indels 4; Gaps 4;
Qy 1 GCTAGCTTAAGTAAGCCCATTTTGCNAGCGCATGGGAAAAATACATTAAGTACAGAAATAGAGA 60
Db 2298 GCTAGCTTAAGTAAGCCCA-TTTGCAAGGCAT-TTGAAAAATACATAACTAGAGAAATAGAGA 2355
Qy 61 AGTTCAGATCAAGTCAGAGCAAGAACAGAGAGAAATATGGCCAAACAGGATATCTGT 120
Db 2356 AGTTCAGATCAAGTCAGAGCAAGAACAGAGAGAAATATGGCCAAACAGGATATCTGT 2415
Qy 121 GCTAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGAAACAGGAGAAAT-TGGGCCA 179
Db 2416 GCTAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGAAACAGGAGAAATATGGGCCA 2475
Qy 180 AACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGTCCCC 239
Db 2476 AACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGTCCCC 2535
Y 240 AGATCGGTCGCGCCCTCAGCAGTTCCTAGTA 270
Db 2536 AGAT-CGGTCCCGCTCAGCAGTTCCTAGTA 2565
RESULT 12
S66424/c
LOCUS S66424 544 bp DNA linear ROD 06-JAN-1994
DEFINITION c-myc [5' LTR, provirus insertion] [mice, plasmacytoma RFPc 2782,
Genomic Mutant, 544 nt].
ACCESSION S66424
VERSION S66424.1 GI:439771
KEYWORDS
SOURCE Mus sp. plasmacytoma RFPc 2782.
ORGANISM Mus sp.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 544)
AUTHORS Shaughnessy, J.D. Jr., Owens, J.D. Jr., Wiener, F., Hilbert, D.M.,
Huppi, K., Potter, M. and Mushinski, J.F.
TITLE Retroviral enhancer insertion 5' of c-myc in two
translocation-negative mouse plasmacytomas upregulates c-myc
expression to different extents
JOURNAL Oncogene 8 (11), 3111-3121 (1993)
MEDLINE 94020837
PUBMED 8414513
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 138854] from the original journal article.

This sequence comes from Fig. 4b.
FEATURES
source Location/Qualifiers
1. .544
/organism="Mus sp."
/db_xref="taxon:10095"
1. .544
/gene="c-myc"
BASE COUNT 121 a 129 c 117 g 177 t
ORIGIN
Query Match 77.4%; Score 209; DB 10; Length 544;
Best Local Similarity 94.9%; Pred. No. 2.4e-54;
Matches 259; Conservative 0; Mismatches 10; Indels 4; Gaps 4;
Qy 1 GCTAGCTTAAGTAAGCCCATTTTGCNAGCGCATGGGAAAAATACATAACTAGAGAAATAGAGA 60
Db 392 GCTAGCTTAAGTAAGCCCATTTTGCNAGCGCAT-TGAAAAATAGATAACTAGAGAAATAGAGA 334
Qy 61 AGTTCAGATCAAGTCAGAGCAAGAACAGAGAGAAATATGGCCAAACAGGATATCTGT 120
Db 333 AGTTCAGATCAAGTCAGAGCAAGAACAGAGAGAAATATGGCCAAACAGGATATCTGT 274
Qy 121 GGTAAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGAAACAGGAGAAAT-TGGGCC 178
Db 273 GGTAAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGAAACAGGAGAAATATGGGCC 214
Qy 179 AACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGTCC 237
Db 213 AACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCCCAAGAACAGTGTGGTCC 154
Qy 238 CCAGATCGGTCGCGCCCTCAGCAGTTCCTAGTA 270
Db 153 CCAGATCGGTCGCGCCCTCAGCAGTTCCTAGTA 121
RESULT 13
AX166278
LOCUS AX166278 4148 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 1 from Patent WO0137881.
ACCESSION AX166278
VERSION AX166278.1 GI:14546734
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 4148)
AUTHORS von Laer, M.D.
TITLE Gene therapy of hiv-positive patients by the expression of
membrane-anchored gp41 peptides
JOURNAL Patent: WO 0137881-A 1 31-MAY-2001;
Heinrich-pette-Institut (DE)
FEATURES
source Location/Qualifiers
1. .4148
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="M87 (STHM) - Retroviral Vektor MPIN, der 5' von
der IRES-NEO-Kassette ein Insert enthaelt, das fuer
Membran-verankertes T-20-PEPTIDEkodiert"
31. .36
/note="NheI-Schnittstelle"
misc_feature 370. .377
misc_feature /note="AscI-Schnittstelle"
misc_feature 400. .405
misc_feature /note="XmaI-Schnittstelle"
misc_feature 696. .701
misc_feature /note="SacII-Schnittstelle"
misc_feature 1421. .1426
misc_feature /note="SacII-Schnittstelle"
misc_feature 1429. .1436
misc_feature /note="NotI-Schnittstelle"
CDS 1438. .1773
/note="Bereich im M87 (STHM)-Insert, der fuer
Membran-verankertes T-20-Fusionsprotein einschliesslich

promoter is 993"
/citation=[3]
1061.1076
/note="Multiple cloning site (MCS): contains the unique
restriction sites XhoI, SacII, NotI"
1160.1224
/note="SV40 small T antigen intron provides splice signal
for inserts cloned into the MCS"
1860.1865
/note="The polyA signal from the SV40 early region
provides polyadenylation of inserts cloned in the MCS"
1881
1931.19875
/note="complete genome of bovine papilloma virus type 1;
The BPV genome provides for episomal maintenance of the
vector pBPV when the vector is transfected into mammalian
cells"
/citation=[5]
complement(10647)
/note="Plasmid origin of replication for replication in E.
coli. base 10647 represents the first base of the newly
synthesized strand"
/direction=left
complement(11406.12336)
/gene="bla"
complement(11406.12266)
/gene="bla"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AA57075.1"
/db_xref="GI:595689"
/translation="MSIQHFRVALIPFFAFLPVPFHPETLVKVKDAEDOLGARVGY
IEDLSKGLTLESFRPEKPMSTKVLKGLGAVLSRDVAGQEOGRIHYSONDLVE
YSPVTEKHGTDGTVRELCSAATMSDNTAANLLITIGSPKELTAPLHNMGDRVREL
DRNEPELNEAIPNDERDTTPAMATTLRLKLTIGELTILASRQQLIDWMEKRVAGPL
LRSAIPAGWFIADKSGAGERGSGIILAAIGPDGPKSRIVVIYITGSAITMDERNRQIA
EIGASLIKHW"
complement(12308.12336)
/gene="bla"
promoter
3381 a 2852 c 3005 g 3278 t
BASE COUNT
ORIGIN
Query Match 94.3%; Score 254.6; DB 12; Length 12516;
Best Local Similarity 96.7%; Pred. No. 2e-68;
Matches 260; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GCTAGCTTAAGTAACGCCATTTTCAGGCATGGGAAATAACATACTGAGATAGAGA 60
Db 269 GCTAGCTTAAGTAACGCCATTTTCAGGCATGGGAAATAACATACTGAGATAGAGA 210
61 AGTTCAGATCAAGTTCAGGAACAGAGAAACAGAGATATGGGCCAAACAGAGATATCTGT 120
60 AGTTCAGATCAAGTTCAGGAACAGAGAAACAGAGATATGGGCCAAACAGAGATATCTGT 150
QY 121 GGTAAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGTGTGGAACAGGAGAAATTTGGGCCAA 180
Db 149 GGTAAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGATGAACAGCTGAATTTGGGCCAA 90
QY 181 ACAGGATATCTGTGTAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGATGTTGCCCA 240
Db 89 ACAGGATATCTGTGTAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGATGTTGCCCA 30
QY 241 GATCGCGTCCCGCTCAGCAGTTCCTAG 269
Db 29 GATCGCGTCCCGCTCAGCAGTTCCTAG 1
RESULT 10
AX449148
LOCUS
DEFINITION
ACCESSION
AX449148
Sequence 2 from Patent WO0238782.
AX449148
6575 bp
DNA
linear
PAT 03-JUL-2002

AX449148.1 GI:21697950
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 6575
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Plasmid pAIM EIB55K, das eine Expressionskassette
fuer EIB55K enth lt."
BASE COUNT 1623 a 1566 c 1695 g 1691 t
ORIGIN
Query Match 91.1%; Score 246; DB 6; Length 6575;
Best Local Similarity 99.3%; Pred. No. 9.1e-66;
Matches 268; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 GCTAGCTTAAGTAACGCCATTTTCAGGCATGGGAAATAACATACTGAGATAGAGA 60
Db 432 GCTAGCTTAAGTAACGCCA-TTTGCAAGCAT-GGAAATAACATACTGAGATAGAGA 489
QY 61 AGTTCAGATCAAGTTCAGGAACAGAGAAACAGAGATATGGGCCAAACAGAGATATCTGT 120
Db 490 AGTTCAGATCAAGTTCAGGAACAGAGAAACAGAGATATGGGCCAAACAGAGATATCTGT 549
QY 121 GGTAAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGTGTGGAACAGGAGAAATTTGGGCCAA 180
Db 550 GGTAAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGTGTGGAACAGGAGAAATTTGGGCCAA 609
QY 181 ACAGGATATCTGTGTAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGATGTTGCCCA 240
Db 610 ACAGGATATCTGTGTAAGCAGTTCCTCCCGCTCAGGCCCAAGAACAGATGTTGCCCA 669
QY 241 GATCGCGTCCCGCTCAGCAGTTCCTAG 270
Db 670 GATCGCGTCCCGCTCAGCAGTTCCTAG 699
RESULT 11
MSYMS
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
MSYMS
2854 bp ss-RNA
linear
VRL 02-AUG-1993
Myeloproliferative sarcoma virus proviral v-mos gene, 3' LTR.
K01683
GI:332208
c-myc proto-oncogene; coat protein; envelope protein;
envelope-associated protein; mos oncogene.
Myeloproliferative sarcoma virus proviral DNA.
Myeloproliferative sarcoma virus
viruses; Retroviral viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.
1 (bases 1 to 2854)
Stacey,A., Arbutnot,C., Kollek,R., Coggin,L. and Ostertag,W.
Comparison of myeloproliferative sarcoma virus with Moloney murine
sarcoma virus variants by nucleotide sequencing and heteroduplex
analysis
J. Virol. 50 (3), 725-732 (1984)
84216451
6328002
[1] reports an amber mutation (deletion of one 'c') in the envelope
gene of the MPSV genome, at the junction of the envelope/mos genes.
What would normally be the twentieth codon is now a termination
codon. At base 973 an in-phase start codon was found. The MPSV
sequence is compared with two temperature sensitive mutants derived
from it, Mo-MuSV variant M1 and Moloney murine leukemia virus
(Mo-MuLV) by heteroduplex mapping. MPSV wild-type contained a 1 kb
deletion from the pol gene. It also contained sequences related to
v-mos genes.

```

|||||
Db 289 GCTAGCTTAAGTAAAGCCATTTTCAAGGCATGGGAAAAATACATACTGAGAAATAGAGA 230
|||||
Qy 61 AGTTACATCAAGTCAAGGAAACAGAGAGAGAGATATGGCCAAACAGGATATCTGT 120
|||||
Db 229 AGTTACATCAAGTCAAGGAAACAGAGAGAGAGATATGGCCAAACAGGATATCTGT 170
|||||
Qy 121 GGTAAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGTGGAAACAGGAGAAATGGGCCAA 180
|||||
Db 169 GGTAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGTGGAAACAGGAGAAATGGGGCAA 110
|||||
Qy 181 ACAGGATATCTGCTAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 240
|||||
Db 109 ACAGGATATCTGCGTAAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 50
|||||
Qy 241 GATCGGTCCCGCCTCAGCAGTTCTTAG 269
|||||
49 GATCGGTCCCGCCTCAGCAGTTCTTAG 21

RESULT 8
E28258/c
LOCUS E28258 3263 bp DNA linear PAT 18-JUN-2001
DEFINITION Purification of higher transcription complex from nonhuman
transgenic animal.
ACCESSION E28258
VERSION E28258.1 GI:13025292
KEYWORDS JP 1999004638-A/10.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3263)
AUTHORS Bernd,K., Eric,B., Michael,M. and Greg,P.
TITLE Purification of higher transcription complex from nonhuman
transgenic animal
JOURNAL Patent: JP 1999004638-A 10 12-JAN-1999;
HOECHST AG
COMMENT OS Unidentified
PN JP 1999004638-A/10
PD 12-JAN-1999
PF 26-MAY-1998 JP 1998144743
PR 26-MAY-1997 DE
PI BERND KIRSCHBAUM,ERIC BERUGURUNTO,MICHAEL MEISTERERST, PI
GREG PORITTSU
PC A01K67/027,C12N15/09,C12P21/02//C12P21/08,(C12N15/09,
C12R1.91), PC
C12N15/00,
PC (C12N15/00,C12R1.91)
CC Strandedness: Single;
Topology: Linear;
FH Key Location/Qualifiers
FT exon 1..3263.
Location/Qualifiers
source
1..3263
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 860 a 794 c 756 g 853 t
ORIGIN

Query Match 94.3%; Score 254.6; DB 6; Length 3263;
Best Local Similarity 96.7%; Pred. No. 1.7e-68;
Matches 260; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 GCTAGCTTAAGTAAAGCCATTTTGAAGGCATGGGAAAAATACATACTGAGAAATAGAGA 60
|||||
Db 289 GCTAGCTTAAGTAAAGCCATTTTGAAGGCATGGGAAAAATACATACTGAGAAATAGAGA 230
|||||
Qy 61 AGTTACATCAAGTCAAGGAAACAGAGAGAGATATGGCCAAACAGGATATCTGT 120
|||||
Db 229 AGTTACATCAAGTCAAGGAAACAGAGAGAGATATGGCCAAACAGGATATCTGT 170
|||||
Qy 121 GGTAAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGTGGAAACAGGAGAAATGGGCCAA 180
|||||
Db 169 GGTAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGTGGAAACAGGAGAAATGGGGCAA 110
|||||
Qy 181 ACAGGATATCTGCTAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 240
|||||
Db 109 ACAGGATATCTGCGTAAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 50
|||||
Qy 241 GATCGGTCCCGCCTCAGCAGTTCTTAG 269
|||||
49 GATCGGTCCCGCCTCAGCAGTTCTTAG 21

```

```

Qy 181 ACAGGATATCTGTTAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 240
|||||
Db 109 ACAGGATATCTGCGTAAAGCAGTTCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 50
|||||
Qy 241 GATCGGTCCCGCCTCAGCAGTTCTTAG 269
|||||
Db 49 GATCGGTCCCGCCTCAGCAGTTCTTAG 21
|||||

RESULT 9
XXU13843/c
LOCUS XXU13843 12516 bp DNA
DEFINITION PBV cloning vector, complete sequence.
ACCESSION U13843
VERSION U13843.1 GI:595688
KEYWORDS bovine papilloma virus; metallothionein I promoter; Moloney sarcoma
virus enhancer; beta-lactamase.
SOURCE unidentified cloning vector.
ORGANISM unidentified cloning vector.
REFERENCE 1 (bases 1 to 12516)
AUTHORS Malone,J.A.
TITLE PBV: An episomally maintained mammalian expression vector based on
bovine papilloma virus
JOURNAL Unpublished (1994)
REFERENCE 2 (bases 1 to 391)
AUTHORS Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
TITLE Nucleotide sequences of integrated Moloney sarcoma provirus long
terminal repeats and their host and viral junctions
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
MEDLINE 81054687
PUBMED 6254003
REFERENCE 3 (bases 692 to 1060)
AUTHORS Glanville,N., Durnam,D.M. and Palmiter,R.D.
TITLE Structure of mouse metallothionein-I gene and its mRNA
JOURNAL Nature 292 (5820), 267-269 (1981)
MEDLINE 81245168
PUBMED 7254320
REFERENCE 4 (bases 939 to 949)
AUTHORS Searle,P.F., Stuart,G.W. and Palmiter,R.D.
TITLE Building a metal-responsive promoter with synthetic regulatory
elements
JOURNAL Mol. and Cell. Biol. 5, 1480-1489 (1985)
REFERENCE 5 (bases 1931 to 9875)
AUTHORS Chen,E.Y., Howley,P.M., Levinson,A.D. and Seeburg,P.H.
TITLE The primary structure and genetic organization of the bovine
papillomavirus type 1 genome
JOURNAL Nature 299 (5883), 529-534 (1982)
MEDLINE 83012974
PUBMED 6289124
REFERENCE 6 (bases 1 to 12516)
AUTHORS Malone,J.A.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1994) James A. Malone, International Technical
Services, Molecular Biology Reagents Division, Pharmacia Biotech
Inc., 2202 N. Bartlett Ave., Milwaukee, WI 53202-1009, USA

FEATURES
source
1..12516
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
/lab_host="Escherichia coli"
complement(1..391)
/standard_name="Moloney Sarcoma Virus (MSV) enhancer"
/citation=[2]
939..949
/standard_name="metal response element (MRE)"
/citation=[4]
965..970
/gene="mouse metallothionein I"
965..970
/gene="mouse metallothionein I"
/note="The first base of nascent transcripts from the MT-I
misc_signal
enhancer
gene
TATA_signal

```

MSV (Moloney murine sarcoma virus) from transformed mink cells, clone P-600-3.
Moloney murine leukemia virus
Viruses; Retroviruses; Retroviridae; Gammaretrovirus.
1 (bases 1 to 903)
Dhar.R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions
Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
81054687
ON Oct 4, 1994 this sequence version replaced gi:331983.
[1]: the first and last 11 nucleotides of the terminal repeat sequence (trs) are inverted with respect to each other, and the same four-nucleotide host sequence is found bracketing integrated msv. Bases 233-820 represent the 588 bp left trs. See mmsvtr52 and mmsvtr53 for the following segments. See also mmsvpro for the complete proviral sequence.
EMBL features not translated to GenBank features:
key from to description
CELL <1 232 cellular DNA (from mink)
PROVRL 233 >903 proviral DNA
SITE 233 820 terminal repeat
SITE 229 232 repeated again at right end of provirus.

Location/Qualifiers
1..903
/organism="Moloney murine leukemia virus"
/proviral
/db_xref="taxon:11801"
238 a 245 c 215 g 205 t
352 bases upstream of pVIIII site.

Query Match 94.7%; Score 255.6; DB 14; Length 903;
Best Local Similarity 96.7%; Pred. No. 7.2e-69;
Matches 261; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY	1	GCTAGCTTAAGTACGGCATTTC	CAGGCGATGGGAAATACTA	CTGAGAATAGAGA	60
Db	261	GCTAGCTTAAGTACGGCATTTC	CAAGGCGATGGGAAATACTA	CTGAGAATAGAGA	320
QY	61	AGTTCAGATCAAGTCAAGGAAC	ACAGAGAGATATGGGCCA	ACAGAGATATCTGT	120
Db	321	AGTTCAGATCAAGTCAAGGAAC	ACAGAGAGATATGGGCCA	ACAGAGATATCTGT	380
QY	121	GGTAAGCAGTTCCTGCCCGCT	CAGGCGCAAGACAGTTG	GAACAGGAGATTTGGGCCAA	180
Db	381	GGTAAGCAGTTCCTGCCCGCT	CAGGCGCAAGACAGTTG	GAACAGGAGATTTGGGCCAA	440
QY	181	ACAGGATATCTGTGAAGCAGTTC	TGCGCCCGCTCAGGCGCA	ACAGAGATGTCGCCCA	240
Db	441	ACAGGATATCTGTGAAGCAGTTC	TGCGCCCGCTCAGGCGCA	ACAGAGATGTCGCCCA	500
QY	241	GATCGGCTCCCGCCTCAGCAGTTC	TCTAGA	270	
Db	501	GATCGGCTCCCGCCTCAGCAGTTC	TCTAGA	530	

RESULT 6
REMSV3
LOCUS
DEFINITION
Left long terminal repeat of Moloney Murine sarcoma virus. (as integrated into mink DNA).
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

REMSV3
903 bp DNA linear VRL 04-APR-2000
Left long terminal repeat of Moloney Murine sarcoma virus. (as integrated into mink DNA).
V01182 J02268
V01182.1 GI:61640
terminal repeat.
Moloney murine sarcoma virus.
Moloney murine sarcoma virus.
Viruses; Retroviruses; Retroviridae; Gammaretrovirus.
1 (bases 1 to 903)
Dhar.R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions

JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	77 (7)	3937-3941	(1980)
MEDLINE	81054687			
PUBMED	6254003			
FEATURES	Location/Qualifiers			
source	233..903			
	/organism="Moloney murine sarcoma virus"			
	/proviral			
	/db_xref="taxon:11809"			
	/focus			
source	1..232			
	/organism="Mustela vison"			
	/db_xref="taxon:9667"			
misc_feature	229..232			
	/note="repeated again at right end of provirus"			
misc_feature	233..820			
	/note="terminal repeat"			
BASE COUNT	238 a 245 c 215 g 205 t			
ORIGIN				
Query Match	94.7%;	Score 255.6;	DB 14;	Length 903;
Best Local Similarity	96.7%;	Pred. No. 7.2e-69;		
Matches 261;	Conservative	0;	Mismatches 9;	Indels 0; Gaps 0;
QY	1	GCTAGCTTAAGTAACGCCATTTC	CAAGGCCATGGGAAAAATACATACTGAGAATAGAGA	60
Db	261	GCTAGCTTAAGTAACGCCATTTC	CAAGGCCATGGGAAAAATACATACTGAGAATAGAGA	320
QY	61	AGTTCAGATCAAGTCCAGGACAGAGAACAGAGAGAATATGGGCCAACACGAGTATCTGT	120	
Db	321	AGTTCAGATCAAGTCCAGGACAGAGAACAGAGAGAATATGGGCCAACACGAGTATCTGT	380	
QY	121	GGTAAGCAGTTCCTGCCCGCTCAGGGCCAAAGACAGTGGGAACAGGAGAATTTGGGCCAA	180	
Db	381	GGTAAGCAGTTCCTGCCCGCTCAGGGCCAAAGACAGTGGGAACAGGAGAATTTGGGCCAA	440	
QY	181	ACAGGATATCTGCTAAGCAGTTCCTGCCCGCTCAGGGCCAAAGACAGATGCTCCCA	240	
Db	441	ACAGGATATCTGCTAAGCAGTTCCTGCCCGCTCAGGGCCAAAGACAGATGCTCCCA	500	
QY	241	GATCGGTCCCGCCTCAGCAGTTTCTAGA	270	
Db	501	GATCGGTCCCGCCTCAGCAGTTTCTAGA	530	
RESULT 7				
AX002278/c				
LOCUS	AX002278	3263 bp	DNA	linear PAT 10-MAR-2000
DEFINITION	Sequence 15 from Patent EP0881288.			
ACCESSION	AX002278			
VERSION	AX002278.1	GI:7241968		
KEYWORDS	unidentified.			
SOURCE	unidentified			
ORGANISM	unclassified.			
REFERENCE	1 (bases 1 to 3263)			
AUTHORS	Berglund, E.D. and Kirschbaum, B.D.			
TITLE	Purification of higher order transcription complexes from transgenic non-human animals			
JOURNAL	Patent: EP 0881288-A 15 02-DEC-1998;			
HOECHST AG (DE)				
FEATURES	Location/Qualifiers			
source	1..3263			
	/organism="unidentified"			
	/db_xref="taxon:32644"			
exon	1..3263			
BASE COUNT	860 a 794 c 756 g 853 t			
ORIGIN				
Query Match	94.3%;	Score 254.6;	DB 6;	Length 3263;
Best Local Similarity	96.7%;	Pred. No. 1.7e-68;		
Matches 260;	Conservative	0;	Mismatches 9;	Indels 0; Gaps 0;
QY	1	GCTAGCTTAAGTAACGCCATTTC	CAAGGCCATGGGAAAAATACATACTGAGAATAGAGA	60

```
APKLLIYDTSNLASGSPSRFSGSGSDTYTETISSLOPEDIATYYCQWSSHIPTFGG
GTRVEIK"
1434..1595
<1596..1916
/codon_start=1
/product="LIGHT CHAIN HMAB425, CONSTANT REGION"
/protein_id="CAB42568.1"
/db_xref="GI:4774644"
/translation="TVAAPSVFIIPPDEQLKSGTASVVCILLNNFYPREAKVOMKVDN
ALQSGNQSVTEQDSKSDSYLSSTLTSLKADYKHKVYACEVTHQGLSSPVTKSFN
RGEC"
CDS
2582..4540
/codon_start=1
/transl_table=1
/product="FUSION PROTEIN: HEAVY CHAIN HMAB425 +
TNFALPHA(FROM 4064)"
/protein_id="CAB42569.1"
/db_xref="GI:4774645"
/translation="MDWTWRFCELLAVAPGAHSQVLVSGAEVKKPGASVKVSKAS
GYFTSHMHWVRQAPGGLEWIEGFNPSNGRTNYNEKFSKATMTVDTSNTAYMEL
SSLRSEDTAVYACRDYDIDYDFYDYGQGLTVTVSSGEWILCAWAQLCPTPRSHGT
TSLAASITKGPSVFLPAPSSKSTSGTAAALGCLVKDYFPEPTVSWNSGALTSGVHTFP
AVLQSLGLSVFLVPPSSSLGTQTYICNVNHRPSNTKVDKVEPKSCDKTHTCPPC
PAPELLGSPVFLPPPKPTLMISRTPEVTVVDVSHEDPEVKENWYVDGVENHNA
KTRPREQYNSTRVSVLVFLVHQLDWLNGKEYCKVSKNALPAIEKTLISAKQGPKE
PQVTLPPSDELTKNOVSLVCLVGLGFPISDAVENSNGQFPENNYKTPTPVLDSDGS
FFLYSKITVDSKRWQGNFVSCSYMHEALHNHYTOKLSLSPGKRVRSSTSPDKPV
AHVYANPQAGQQLWLRNALLANGVELNDQLVVPSEGLYLIYSQVLFKGGQCPDS
THVLLTHITRIAYSYQTKVNLLSAISKPCQRETPEGAEPWPEPIYLGQVFLQLEK
DRLSAEINRPDLDFAESGQVYFGIIL"
misc_RNA
4565..5279
/product="5' UTR FROM POLIVIRUS PLUS IRES PLUS INTRON"
5280..5879
CDS
/codon_start=1
/function="SELECTION MARKER"
/transl_table=1
/product="PUROMYCIN ACETYL TRANSFERASE"
/protein_id="CAB42570.1"
/db_xref="GI:4774646"
/translation="MTEYKPTVRLATRDVPRVAVRTLAFAFADYPATRHRTVDPRHIE
RVTEQLFELTRVLGIDIGKVMVADDAVAWVTPESVEAGAVAEIGPRAELSGSR
LAAQOMQELFLTRVGLKHPKPEAFWFLATGVSPDHQGLGSAVVLGVEAERAGVPAFL
ETSAPRNLPYERLGFYVADVECPKDRATWCMTRKFGA"
misc_RNA
5877..8298
/product="DNA SEQUENCE COMPRISING SV40 POLYA (5929-6181)"
/standard_name="SV40 POLYA"
BASE COUNT 1974 a 2364 c 2134 g 1826 t
ORIGIN
Query Match 100.0%; Score 270; DB 6; Length 8298;
Best Local Similarity 100.0%; Pred. No. 3e-73;
Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTAGCTTAAGTAAAGCCATTGTCAGGCGCATGGGAAAATACATACTGAGAATAGAGA 60
|||||
Db 36 GCTAGCTTAAGTAAAGCCATTGTCAGGCGCATGGGAAAATACATACTGAGAATAGAGA 95
|||||
QY 61 AGTTCAGATCAAGGTTCAGGAACAGAGAGAATATGGGCCAACAGGATATCTGT 120
|||||
Db 96 AGTTCAGATCAAGGTTCAGGAACAGAGAGAATATGGGCCAACAGGATATCTGT 155
|||||
QY 121 GGTAAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGAGAAATGGGCCAA 180
|||||
Db 156 GGTAAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGAGAAATGGGCCAA 215
|||||
QY 181 ACAGGATATCTGTGGTAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 240
|||||
Db 216 ACAGGATATCTGTGGTAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 275
|||||
QY 241 GATCGGTCCCGCCCTCAGCAGTTCCTAG 270
|||||
Db 276 GATCGGTCCCGCCCTCAGCAGTTCCTAG 305
```

```
RESULT 4
LOCUS REMSV4
DEFINITION Moloney mouse sarcoma virus right long terminal repeat (as
integrated into mink DNA).
ACCESSION V01183
VERSION V01183.1 GI:61641
KEYWORDS terminal repeat.
SOURCE Moloney murine sarcoma virus.
ORGANISM Moloney murine sarcoma virus
VIRUSES: Retrovirus; Retroviridae; Gammaretrovirus.
REFERENCE 1 (bases 1 to 646)
AUTHORS Dhar,R., McClements,W.L., Enquist,L.W. and Vande Woude,G.F.
TITLE Nucleotide sequences of integrated Moloney sarcoma provirus long
terminal repeats and their host and viral junctions
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3937-3941 (1980)
MEDLINE 81054687
PUBMED 6254003
FEATURES
Source
Location/Qualifiers
1..646
/organism="Moloney murine sarcoma virus"
/db_xref="taxon:11809"
/focus="proviral DNA integrated into Mink"
1..646
/organism="Mustela vison"
/db_xref="taxon:9667"
595..646
/organism="Mustela vison"
/db_xref="taxon:9667"
/organism="cellular DNA"
7..594
/feature="right terminal repeat"
595..598
/feature="repeated before left end of provirus"
BASE COUNT 162 a 174 c 164 g 146 t
ORIGIN
Query Match 94.7%; Score 255.6; DB 14; Length 646;
Best Local Similarity 96.7%; Pred. No. 6.9e-69;
Matches 261; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 GGTAGCTTAAGTAAAGCCATTGTCAGGCGCATGGGAAAATACATACTGAGAATAGAGA 60
|||||
Db 35 GGTAGCTTAAGTAAAGCCATTGTCAGGCGCATGGGAAAATACATACTGAGAATAGAGA 94
|||||
QY 61 AGTTCAGATCAAGGTTCAGGAACAGAGAGAATATGGGCCAACAGGATATCTGT 120
|||||
Db 95 AGTTCAGATCAAGGTTCAGGAACAGAGAGAATATGGGCCAACAGGATATCTGT 154
|||||
QY 121 GGTAAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGAGAAATGGGCCAA 180
|||||
Db 155 GGTAAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGAGAAATGGGCCAA 214
|||||
QY 181 ACAGGATATCTGTGGTAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 240
|||||
Db 215 ACAGGATATCTGTGGTAGCAGTTCCTGCGCCGCTCAGGCGCAAGAACAGAGATGGTCCCA 274
|||||
QY 241 GATCGGTCCCGCCCTCAGCAGTTCCTAG 270
|||||
Db 275 GATCGGTCCCGCCCTCAGCAGTTCCTAG 304
|||||
RESULT 5
LOCUS MLMTRS01
DEFINITION Moloney murine sarcoma provirus LTR and host sequence, segment 1.
ACCESSION J02268
VERSION J02268.1 GI:555246
KEYWORDS terminal repeat.
SEGMENT 1 of 3
```

JOURNAL Patent: WO 0065074-A 1 02-NOV-2000;
 Gesellschaft für Biotechnologische Forschung mbH (GBF) (; DE)
 FEATURES
 source
 1. .270
 /organism="Myceloproliferative sarcoma virus"
 /db_xref="taxon:11813"
 BASE COUNT 85 a 61 c 74 g 50 t
 ORIGIN
 Query Match 100.0%; Score 270; DB 6; Length 270;
 Best Local Similarity 100.0%; Pred. No. 2e-73;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCTAGCTTAAGTAACCCATTTTGCAGGCGATGGGAAAATACATACTAGATAGAGA 60
 Db 1 GCTAGCTTAAGTAACCCATTTTGCAGGCGATGGGAAAATACATACTAGATAGAGA 60
 QY 61 AGTTTCAGATCAAGTTCAGGAACAGAGGAGGAATATGGGCCAAACAGGATATCTGT 120
 Db 61 AGTTTCAGATCAAGTTCAGGAACAGAGGAGGAATATGGGCCAAACAGGATATCTGT 120
 QY 121 GGTAAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 180
 Db 121 GGTAAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 180
 QY 181 ACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 240
 Db 181 ACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 240
 QY 241 GATCGCGTCCCGCCCTCAGCAGTTTCTAGA 270
 Db 241 GATCGCGTCCCGCCCTCAGCAGTTTCTAGA 270
 RESULT 2
 PMBC2TD 3123 bp DNA linear SYN 23-JAN-1995
 LOCUS Plasmid pmBC-2T (expression plasmid-eukaryotic-in vitro) DNA.
 DEFINITION
 X77750.1 GI:456203
 VERSION expression plasmid; Plasmid.
 KEYWORDS synthetic construct.
 SOURCE synthetic construct.
 ORGANISM
 artificial sequences.
 REFERENCE
 1 (bases 1 to 3123)
 Dirks, W., Schaper, F. and Hauser, H.
 AUTHORS A new hybrid promoter directs transcription at identical starts
 TITLE points in mammalian cells and in vitro
 JOURNAL Gene 49, 389-390 (1994)
 REFERENCE 2 (bases 1 to 3123)
 Schaper, F.
 AUTHORS Direct Submission
 TITLE Submitted (14-FEB-1994) F. Schaper, Gesellschaft für
 JOURNAL Biotechnologische Fors., (GBF), Mascheroder Weg 1, 38106
 Braunschweig, FRG
 FEATURES
 source
 1. .3123
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /plasmid="pmBC-2T"
 promoter 5. .455
 /note="from MT7"
 gene 463. .521
 /gene="SV40 19S mRNA"
 intron 463. .521
 /gene="SV40 19S mRNA"
 /note="from VP1"
 misc_feature 529. .591
 /note="polylinker site from pUC8"
 polyA_signal 626. .862
 /note="SV40 of VP1"
 terminator 1022. .1157
 /note="from T7"

BASE COUNT 815 a 753 c 761 g 794 t
 ORIGIN
 Query Match 100.0%; Score 270; DB 12; Length 3123;
 Best Local Similarity 100.0%; Pred. No. 2.7e-73;
 Matches 270; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GCTAGCTTAAGTAACCCATTTTGCAGGCGATGGGAAAATACATACTAGATAGAGA 60
 Db 36 GCTAGCTTAAGTAACCCATTTTGCAGGCGATGGGAAAATACATACTAGATAGAGA 95
 QY 61 AGTTTCAGATCAAGTTCAGGAACAGAGGAGGAATATGGGCCAAACAGGATATCTGT 120
 Db 96 AGTTTCAGATCAAGTTCAGGAACAGAGGAGGAATATGGGCCAAACAGGATATCTGT 155
 QY 121 GGTAAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 180
 Db 156 GGTAAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 215
 QY 181 ACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 240
 Db 216 ACAGGATATCTGTGTAAGCAGTTCCTGCGCCGCTCAGGCGCCAAAGAACAGTTCAGGAGGAGGATATCTGT 275
 QY 241 GATCGCGTCCCGCCCTCAGCAGTTTCTAGA 270
 Db 276 GATCGCGTCCCGCCCTCAGCAGTTTCTAGA 305
 RESULT 3
 A70359 8298 bp DNA circular PAT 07-MAY-1999
 LOCUS Sequence 1 from Patent WO9811241.
 DEFINITION
 A70359
 ACCESSION A70359.1 GI:4774641
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 1 (bases 1 to 8298)
 Von, H.I., Bruemmer, W., Burge, C., Riske, E., Dunker, R., Welge, T.,
 AUTHORS Hauser, H. and Mielke, C.
 TITLE OLIGOCISTRONIC EXPRESSION SYSTEM FOR THE PRODUCTION OF HETEROMERIC
 PROTEINS
 JOURNAL Patent: WO 9811241-A 1 19-MAR-1998;
 VON HOEEN IKA (BE)
 FEATURES
 source
 1. .8298
 /organism="Escherichia coli"
 /strain="K12"
 /db_xref="taxon:562"
 /cell_line="BHK-21"
 /cell_type="FIBROBLAST"
 1. .904
 promoter 905. .976
 intron 977. .1018
 CDS
 /codon_start=1
 /transl_table=11
 /product="LEADER SEQUENCE (PART)"
 /protein_id="CAB42566.1"
 /db_xref="GI:4774642"
 /translation="MGWSCILFLVATA"
 1019. .1106
 /function="5'UTR POLIOVIRUS"
 <1107. .>1433
 /function="FRS, CDRS"
 /codon_start=1
 /transl_table=11
 /product="LIGHT CHAIN HMAB425, VARIABLE REGION, PLUS
 LEADER(REST)"
 /protein_id="CAB42567.1"
 /db_xref="GI:4774643"
 /translation="VHSDIQMTQSPSSLSASVGRVITCSASSSVTYMYWYQKPGK"

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March. 5, 2003, 22:29:16 ; Search time 1487.61 Seconds
(without alignments)
5282.140 Million cell updates/sec

Title: US-09-980-548-1
Perfect score: 270
Sequence: 1 gctagcttaagtaacgccat.....cgccctcagcagttcttaga 270

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	270	100.0	270	6	AX040909 Sequence
2	270	100.0	3123	12	PMBC2TD
3	270	100.0	8298	6	A70359 Sequence 1
4	255.6	94.7	646	14	REMSV4
5	255.6	94.7	903	14	MLMTR01
6	255.6	94.7	903	14	REMSV3
7	254.6	94.3	3263	6	AX002278 Sequence
8	254.6	94.3	3263	6	E28258
9	254.6	94.3	12516	12	XX013843
10	246	91.1	6575	6	AX449148
11	223	82.6	2854	14	MSYMO5
12	209	77.4	544	10	S66424
13	208.2	77.1	4148	6	AX166278 Sequence
14	208.2	77.1	9318	6	AR028672 Sequence
15	206.4	76.4	733	14	MLM3LTR
16	206	76.3	5893	14	REMLV
17	206	76.3	5894	14	AF033812
18	206	76.3	5894	14	MLAPRO
19	202.8	75.1	508	6	AX107878 Sequence
20	202.8	75.1	594	14	MMSLTR1
21	202.8	75.1	635	6	AR016499 Sequence
22	202.8	75.1	635	6	AR096882 Sequence
23	202.8	75.1	636	6	E05956 Promoter/en
24	202.8	75.1	702	6	I77226 Sequence 16
25	202.8	75.1	1200	10	RATDLTR
26	202.8	75.1	1860	10	AX133244S3
27	202.8	75.1	3097	6	AX107931 Sequence
28	202.8	75.1	4207	6	AX359930 Sequence
29	202.8	75.1	4207	6	AX382144 Sequence
30	202.8	75.1	4210	6	AX359931 Sequence
31	202.8	75.1	4210	6	AX382145 Sequence
32	202.8	75.1	4273	6	AR112508
33	202.8	75.1	4273	6	AR201984
34	202.8	75.1	4273	12	PLTRPOLY
35	202.8	75.1	4661	6	AX359936 Sequence
36	202.8	75.1	4661	6	AX382150 Sequence
37	202.8	75.1	4776	6	A95117 Sequence 1
38	202.8	75.1	4776	6	A95152 Sequence 1
39	202.8	75.1	4922	6	AX146812 Sequence
40	202.8	75.1	5109	6	I56770 Sequence 1
41	202.8	75.1	5117	12	AF132210 Cloning v
42	202.8	75.1	5130	6	AX359935 Sequence
43	202.8	75.1	5130	6	AX382149 Sequence
44	202.8	75.1	5176	6	I56771 Sequence 2
45	202.8	75.1	5177	6	AX018983 Sequence

ALIGNMENTS

RESULT 1
AX040909
LOCUS AX040909 270 bp DNA linear PAT 23-NOV-2000
DEFINITION Sequence 1 from Patent WO0065074.
ACCESSION AX040909
VERSION AX040909.1 GI:11340531
KEYWORDS
SOURCE Myeloproliferative sarcoma virus.
ORGANISM Myeloproliferative sarcoma virus
Viruses; Retroviral viruses; Retroviridae; Mammalian type C
retroviruses; 1-Mammalian type C virus group.
REFERENCE 1 (bases 1 to 270)
AUTHORS Mueller, P., Geserick, C., Schroeder, K. and Hauser, H.
TITLE Promoter-transactivator system for inducible high-level mammalian gene expression with the option of cell growth control